

PAST AS PROLOGUE

The National Academy of Education at 50

Members Reflect



PAST AS PROLOGUE
The National Academy of Education at 50
Members Reflect

Michael J. Feuer, Amy I. Berman, and Richard C. Atkinson, *Editors*

National Academy of Education
Washington, DC

NATIONAL ACADEMY OF EDUCATION 500 Fifth Street, NW Washington, DC 20001

International Standard Book Number: 978-0-9969495-0-7

Library of Congress Control Number: 2015955319

Additional copies of this report are available from the National Academy of Education, 500 Fifth Street, NW, Washington, DC 20001; <http://www.naeducation.org>.

Copyright 2015 by the National Academy of Education. All rights reserved.

Printed in the United States of America

Suggested citation: Feuer, M. J., Berman, A. I., & Atkinson, R. C. (Eds.). (2015). *Past as Prologue: The National Academy of Education at 50. Members Reflect*. Washington, DC: National Academy of Education.

NATIONAL ACADEMY *of* EDUCATION

The **National Academy of Education** advances high quality education research and its use in policy formation and practice. Founded in 1965, the Academy consists of U.S. members and foreign associates who are elected on the basis of outstanding scholarship related to education. Since its establishment, the Academy has undertaken research studies that address pressing issues in education, which are typically conducted by members and other scholars with relevant expertise. In addition, the Academy sponsors professional development fellowship programs that contribute to the preparation of the next generation of scholars.

Foreword

Fifty years ago American education changed. Congress enacted critical and wide-reaching federal laws, the executive branch demanded the enforcement of revolutionary Supreme Court mandates, and we had finally come to terms with the abomination of racial injustice by passing the Civil Rights Act. More than ever in our history, Washington inserted itself in what was until then a state and local education system. To appreciate the zeitgeist, listen to the words of President Lyndon B. Johnson, as he signed the Elementary and Secondary Education Act: He said the law will “bridge the gap between helplessness and hope for [this country’s] educationally deprived children ... no law I have signed or will ever sign means more to the future of America.” And there was more to come: 6 months later, Congress enacted the Higher Education Act (HEA), swinging open the doors of college to all students. The effects? In 1965, less than 10 percent of the population 25 years and older had attended 4 or more years of college. By 2014, that number was close to 34 percent. But there is still more: the year 1965 also saw the introduction of Head Start, the federal government’s foray into early childhood education, which made a national statement about the imperative to break the cycle of poverty through comprehensive programming and preschool to poor children.

At the risk of confusing cause and effect, these changes were happening just as giants of education research (James Conant, Lawrence Cremin, Frank Keppel, Israel Scheffler, Ralph Tyler, and others) were discussing the need for better data and enhanced scholarship concerning the improvement of education. And thus the National Academy of Education (NAEd) was established 50 years ago.

It was a radical idea, if not for science and research generally, then certainly for the idea of education research specifically: an Academy “dedicated to the advancement of scholarship concerning the ends and means of education, in all its forms, in the United States and abroad.” Modeled after the National Academy of Sciences, which was founded 102 years earlier, the NAEd would bring together outstanding scholars and provide a formal structure for them to engage in instruction, debate, and research “transcending the bounds of any single institution.” With the generous and enlightened support of the Carnegie Corporation, the vision was realized.

With this anthology we begin a year of celebration of our 50th anniversary, published in time for the annual gathering of members, fellows, and friends in Washington, DC, in October 2015. We asked members to use the occasion to reflect on the Academy’s commitment to addressing the most pressing educational issues of our day, and to think about how we might become even more influential in the years ahead. We are awed and inspired by the words of our colleagues collected here.

Not surprisingly, with the coinciding anniversaries of the NAEd and the laws that fundamentally changed the American educational and social landscape, many of our members critiqued the current state of our educational system, offered suggestions for improvement, and looked to the future role of the NAEd. The 52 articles span a wide range of topics, which we have clustered into 8 sections.

Many submissions address early childhood and K–12 education. Bruce Alberts covers the crucial role of teachers and their inclusion in educational decision-making, Richard Atkinson reminisces on his work about memory and computer-assisted instruction, William Damon addresses the importance of civics education as it relates to the persistent problems of inequality, Diane Slaughter Kotzin addresses “play” in early childhood education, Marshall (Mike) Smith focuses on systemic solutions for the systemic problems facing our educational institutions, Catherine Snow reviews the history of the research on teaching of reading, Noreen Webb addresses student engagement and participation, and Roger Weissberg attends to social and emotional learning.

We include a section titled “Assessment and Measurement in Education,” the importance of which in the history of education policy and practice is abundantly obvious. Our Academy and many of its members have examined the uses (and misuses) of assessment in our nation’s schools. Here Paul Black describes his extensive work in assessment for learning, Edmund Gordon explores assessment to further learning and human development, Marcia Linn argues for the potential of assessments to strengthen online science learning, James Pellegrino explains why assessment systems need to keep pace with the sciences of learning, and Robert

Sternberg describes the failure of assessments to be based on a modern theory of intelligence.

Of the approximately 50 million students in public elementary and secondary schools, less than half are white, 15 percent are black, 26 percent are Hispanic, 5 percent are Asian/Pacific Islander, 1 percent are American Indian/Alaska Native, and 3 percent identify as two or more races. One-tenth of our students are identified as English language learners. Poverty has become more severe: roughly 22 percent of all children live in families with incomes below the federal poverty level and 48 percent of students are eligible for free/reduced-price lunch. In the section "The Education of a Diverse Learning Population," members address the imperative for our schools to provide an equal learning opportunity for all of the nation's students. Greg Duncan and Richard Murnane describe rising income inequality in the United States and what we might do about it, Patricia Gándara retraces the history of language policies, Sonia Nieto laments the dismantling of public education because of segregation and privatization, Alejandro Portes examines the plight and prospects of immigrant students, and Margaret Beale Spencer discusses the continued need to recognize the salience of race.

With the HEA came unprecedented access to the nation's postsecondary system. In 1965, roughly 19 percent of 20–24-year-olds were enrolled in higher education; by 2011, that number more than doubled to 40 percent. Instead of higher education being an ivory tower for the select few, the HEA made it possible for many more students to gain access to higher education. In the section "Challenges in Higher Education," Stephen Ceci examines trends in academic hiring and the promotion of females in academic institutions, Robert Floden takes on recent debates over teacher preparation programs and their evaluation, Michael McPherson explores the history of higher education with an emphasis on social science research contributions in the HEA, Cecilia Rouse examines student completion and affordability, Marta Tienda describes the importance of diversification in the higher education population, and William Tierney discusses the federal government's long involvement in higher education.

A dominant theme in the history of the Academy is the role of research and the quality of evidence it produces. In the section "Modes of Inquiry for Educational Research," the authors comment on the history of educational research and the ongoing debate over evidentiary standards. Margaret Eisenhart, Michael Feuer, Kenji Hakuta, Richard Shavelson, and Maris Vinovskis explore from their vantage points the demand for scientifically valid research, what this demand means, and how to ensure that education research addresses the complexities of our educational system. Michael Cole traces his journey from an experimental learning theorist to a "mushy developmentalist," David Kaplan examines misinterpreta-

tion and misuse of statistical data and the opportunities for the Bayesian paradigm, and James March and Denis Phillips challenge us to broaden our definition of research, with Phillips reminding us that education is an “intrinsic good” and March that “education is a vision, not a production facility.”

Whether and how education research can be “useful” in policy is, of course, an abiding concern of the Academy. This is the main focus of the section called “The Interplay Between Education Research and Policy.” Eric Hanushek argues for more clarity between the roles of scientific education research and education policy, Jack Jennings provides examples of policies that have not been supported by research and the need for good research to reach policymakers, Lorraine McDonnell addresses the need to include an evaluation of the political sustainability of education policies to complement the research on effectiveness, Lauren Resnick addresses the tensions between scholarly research and policy, and Judge David Tatel comments on the important uses of research in judicial decision-making and the hazards of partisanship disguised as research.

These reflections also reveal how the role of the Academy has evolved throughout the past 50 years. Our membership has grown from 20 to almost 200. We extended the scope and depth of our research studies and expanded our role to include providing professional development and mentoring to prepare the next generation of education scholars. The authors whose pieces are included in the section titled “The National Academy of Education: Then, Now, and the Future” comment on the work we have done and what awaits us now. Many of our former NAEd presidents, including Susan Fuhrman, Patricia Graham, Carl Kaestle, Ellen Condliffe Lagemann, Nel Noddings, and Lorrie Shepard provide wise counsel on the challenges faced by the NAEd. Howard Gardner, James Gee, and Susan Moore Johnson provide examples of areas where the NAEd must struggle with internal decisions (such as membership caps) and external concerns for educational scholars (such as encouraging scholars who think on the margins), and Miriam Ben-Peretz provides her unique perspective as a foreign associate.

A major role of the Academy is professional development of the scholarly community. Thanks to the generous sponsorship of the Spencer Foundation, the Academy has helped to improve the research training and career development of 792 postdoctoral fellows and 112 dissertation fellows to date. We have been honored to watch these fellows blossom into the next generation of leading scholars in education, with several later being honored by election to the NAEd membership. This book includes articles from those intimately involved in the selection and training of our fellows, including the recent chair of the professional development committee (Marilyn Cochran-Smith), former chair of the postdoctoral fel-

lowship selection committee (Hilda Borko), and a recent member of the dissertation fellowship selection committee (Lois Weis). We also have the expertise of Adam Gamoran and Larry Hedges, who conducted extensive evaluations examining the effectiveness of the programs. There is a clear consensus here: these programs are vital and with continuing improvement will play a major role in enhancing the research, productivity, and professional development of education scholars in the future.

We are grateful to all our contributors, whose generosity has (again) accumulated in a trove of ideas and suggestions for the field.

— *Michael Feuer*
Amy Berman
Richard Atkinson

Acknowledgments

An extra note of appreciation from Michael Feuer and Dick Atkinson: This effort would not have been imaginable without the incomparable skills and diligence of the National Academy of Education (NAEd) staff. Amy Berman provided world-class editorial and logistical support while continuing her own doctoral studies and other NAEd obligations, for which we are grateful beyond words. The NAEd Executive Director Greg White provided wise and patient advice and assistance, and program officers Abigail Bell and Jack Busbee were always there with their own special expertise. With 52 articles from 54 authors, the work of our colleagues from the National Academies Press—Michael Dudzik, Rachel Marcus, and Estelle Miller—along with copyeditor Nancy Tuveson and graphic designer Judith Brown was spectacular. We are ever so thankful.

Contents

The National Academy of Education: Then, Now, and the Future	1
Being a Foreign Associate of the National Academy of Education: A Globalization Experience <i>Miriam Ben-Peretz</i>	3
The State of the Academy <i>Susan Fuhrman</i>	5
The National Academy of Education: Personal Reflections <i>Howard Gardner</i>	11
The National Academy of Education: The Center, the Margins, and the Future <i>James Paul Gee</i>	15
Presidency of the National Academy of Education: 1985–1989 <i>Patricia Albjerg Graham</i>	19
Membership in the National Academy of Education: An Ongoing Dialogue <i>Susan Moore Johnson</i>	23

Recollections of My Presidency of the National Academy of Education: 1993–1997 <i>Carl Kaestle</i>	29
The National Academy of Education as a Bold Source for Big Ideas <i>Ellen Condliffe Lagemann</i>	35
Thoughts on the National Academy of Education 10 Years Later <i>Nel Noddings</i>	39
If We Know So Much from Research on Learning, Why Are Educational Reforms Not Successful? <i>Lorrie A. Shepard</i>	41
The National Academy of Education and the Professional Development of Education Scholars	53
Celebrating Our Successes and Offering a Cautionary Note: Reflections on the National Academy of Education/Spencer Fellowship Programs <i>Hilda Borko and Lois Weis</i>	55
Professional Development for Early-Career Scholars <i>Marilyn Cochran-Smith</i>	63
Preparing the Next Generation of Education Researchers: Reflections on the Role of the National Academy of Education <i>Adam Gamoran</i>	69
Reflections on the National Academy of Education on Its 50th Anniversary <i>Larry V. Hedges</i>	75
Modes of Inquiry for Educational Research	83
Hooking Up with Romantic Science <i>Michael Cole</i>	85
Legislating the Value of Educational Research <i>Margaret Eisenhart</i>	89
Evidence and Advocacy <i>Michael J. Feuer</i>	95

Office of Educational Research, Institute for Education Sciences, and Shaping Education Research <i>Kenji Hakuta</i>	103
The Future of Quantitative Inquiry in Education: Challenges and Opportunities <i>David Kaplan</i>	109
Education and Don Quixote <i>James G. March</i>	117
Philosophy and the National Academy of Education <i>D. C. Phillips</i>	121
Reflections on <i>Scientific Research in Education</i> <i>Richard J. Shavelson</i>	127
Assessing and Improving Federal Education Research <i>Maris A. Vinovskis</i>	135
Improvement of Early Childhood and K–12 Education	141
Empowering Our Best Teachers: Essential for Producing More Effective Systems of Education in the United States <i>Bruce Alberts</i>	143
Vannevar Sets the Stage <i>Richard C. Atkinson</i>	151
In Search of Broader Visions: An Appreciation and Future Hope for the National Academy of Education <i>William Damon</i>	159
Reflections Upon 50 Years of U.S. Research in Early Childhood, 1962–2012: A Role for the National Academy of Education <i>Diana Slaughter Kotzin</i>	165
Systemic Problems Require Systemic Solutions <i>Marshall S. Smith</i>	167
50 Years of Research on Reading <i>Catherine Snow</i>	175

The Importance of Fostering Productive Student Participation <i>Noreen M. Webb</i>	181
Education to Promote All Students' Social, Emotional, and Academic Competence <i>Roger P. Weissberg</i>	185
The Education of a Diverse Learning Population	193
Income Inequality and American Education <i>Greg J. Duncan and Richard J. Murnane</i>	195
Is U.S. Language Education Policy Outdated? <i>Patricia Gándara</i>	201
Revisiting the High Hopes and Broken Promises of Public Education: Still an Uncertain Future <i>Sonia Nieto</i>	213
Assimilation Without a Blueprint: Children of Immigrants in New Places of Settlement <i>Alejandro Portes</i>	221
Anniversary Reflections as Opportunity Both for Accruing Insights and Acknowledging Needs <i>Margaret Beale Spencer</i>	227
Assessment and Measurement in Education	233
Assessment: Friend or Foe of Pedagogy and Learning <i>Paul Black</i>	235
Conservation and Exploration: What Is, What If, Why Not? <i>Edmund W. Gordon</i>	243
Refining Assessment to Strengthen Online Science Learning <i>Marcia C. Linn</i>	247
Rethinking and Redesigning Educational Assessment <i>James W. Pellegrino</i>	255

Archaic Testing and Teaching in the United States: Why Do They Persist? <i>Robert J. Sternberg</i>	265
Challenges in Higher Education	271
Women in the Academy: Past, Present, and Future <i>Stephen J. Ceci</i>	273
Learning What Research Says About Teacher Preparation <i>Robert E. Floden</i>	279
Federal Involvement in Higher Education Finance <i>Michael McPherson</i>	285
50 Years of Progress: Higher Education, Research, and Reform <i>Cecilia Elena Rouse</i>	291
Affirmative Action and Its Discontents: America's Obsession with Race <i>Marta Tienda</i>	297
The Higher Education Act of 1965: A Half-Century's Worth of Contrasting Public Philosophies and Controversies <i>William G. Tierney</i>	301
The Interplay Between Education Research and Policy	307
Education's Double Helix <i>Eric A. Hanushek</i>	309
Research and Policy: The Need to Tie the Knot <i>Jack Jennings</i>	313
Melding Political Sustainability Analysis with Education Research <i>Lorraine M. McDonnell</i>	317
The National Academy of Education's Role in Bridging Research and Practice <i>Lauren Resnick</i>	325

Education in the Courtroom: The Need for Unbiased Research <i>David S. Tatel</i>	329
National Academy of Education Members	331
Current Members	333
In Memoriam	337

The National Academy of Education: Then, Now, and the Future

Being a Foreign Associate of the National Academy of Education: A Globalization Experience

*Miriam Ben-Peretz*¹

Scholars in all domains have many opportunities for sharing knowledge, experiences, and engagement in joint research and publication. For many years I have engaged with scholars in the United States through conferences, research and joint publications, and teaching in American universities. I am a longstanding member of the American Education Research Association (AERA) and for the past 2 years have been an AERA fellow. In addition, I have twice served on AERA's Committee on International Relations. I have often taught summer courses at American universities, and many of my closest colleagues and friends are American scholars. I have been a foreign associate of the National Academy of Education (NAEd) since 2010. At the NAEd, Marcelo Suarez-Orozco and I organized a joint American-Israeli workshop on the education of immigrant students, held in March 2014. On the basis of this workshop, Jim Banks, Marcelo, and I are editing a book to be published by Teachers College Press and to include papers beyond those presented at the March 2014 workshop.

I believe that the collaboration between American scholars and me, under the auspices of the NAEd, exhibits features that distinguish it from other modes of cooperation among scholars. Working together within the framework of a common organization not only provides scholars with the organization's support but also moves the cooperation from one that con-

¹ Miriam Ben-Peretz is Professor Emerita at the University of Haifa. She was elected to the National Academy of Education in 2010.

siders the backgrounds of two, or more, different research institutions to one that is based on a common ground and context. This common context diminishes consideration of each institutional background and replaces it with membership in a joint enterprise. The special quality of the joint work within the NAEd framework is the expectation of continuation and growth. This ongoing process is different from the usual academic sharing and exchange of knowledge that occurs during scientific conferences, symposia, or meetings.

The NAEd, by including foreign associates in its work, opens its doors to the global world. Such experiences on a worldwide basis offer the promise of a global world of joint scholarly endeavor.

The State of the Academy

Susan Fuhrman¹

In 2008, as part of a Teachers College (TC) trip to visit alumni and donors in Los Angeles, I lunched with Charlotte Cremin, the widow of our former president, Larry. After a very pleasant hour of gossip and updates about TC, Charlotte asked me whether I had ever heard of the National Academy of Education (NAEd). Coincidentally, I had recently been elected president-elect of the NAEd. She proceeded to share her memory of how the NAEd started. Larry, James Conant, and Ralph Tyler were attending the same dinner. In coupled conversations they all agreed to the vision of starting the Academy. She said they never spoke as a group of three that evening, but each left committed to the cause. This story differs a bit from that recorded in the first volume of the *Proceedings of the National Academy of Education* (1965–1974), but it conforms with the depiction of a few leading lights independently reaching the conclusion that an Academy, like the National Academy of Sciences, would be a way to advance educational scholarship. Charlotte’s description of the founders’ goals resonated with my sense of the NAEd’s mission of advancing high-quality research and its use. I felt reinforced that we must build both capacity for education research and demand for that research among policymakers and practitioners.

When I served as President from 2009 to 2013, I came to believe that we are addressing some aspects of that mission extremely well and that

¹ Susan Fuhrman is President of Teachers College, Columbia University. She was elected to the National Academy of Education in 2002.

other aspects remain a challenge for the future and, hopefully, a focus of collective action.

Building Capacity for Education Research

I think we have been extremely successful in building the capacity for research by supporting the development of educational scholars and pushing the boundaries of the field as our understanding of learning and all that influences it expands.

Enhancing capacity was a focus from the Academy's very beginning. It brought together established leaders to span disciplinary boundaries. Interacting with one another, historians, philosophers, psychologists, and other social scientists came to better appreciate the value of multiple perspectives in advancing understanding of educational processes, practices, and systems. We have, of course, continued this work, expanding beyond the "founding disciplines" to include members representing newer fields, such as education policy and teacher quality, and inviting practitioners who have relied on and promoted scholarship in their work.

As the fields we encompass as "education research" grew, so did the topics we studied. Many reports in the 1970s and 1980s centered on educational assessment, reflecting the strong presence of psychologists among members. During my presidency, in contrast, we published *Evaluation of Teacher Preparation Programs: Purposes, Methods, and Policy Options* (Feuer, Floden, Chudowsky, & Ahn, 2013), indicating the growing number of members whose research focuses on teaching, teacher education, and policy.

Strengthening the field also involves making the leadership body, the Academy, reflective of the range of education scholars, not, as in the beginning, the exclusive province of older white males. Diversity could easily be the sole subject of a set of reflections about the Academy: as we strive to focus on gender, racial and ethnic background, and age, we continue to confront the fact that both members and fellows tend to come from a small set of elite institutions. Expanding beyond that group to reach deeper into the Academy is an imperative.

We have also, of course, built capacity through the development of new scholars with our fellowship programs. Thirty years ago the Academy launched the NAEd Spencer Postdoctoral Program. Studies have demonstrated how successful the program has been in advancing the careers of the fellows (Gamoran & Bruch, 2010; Hedges & Hanis, 2005). The mentoring by Academy members and the networking with other young scholars have enabled these beginners to become the leaders of the future, just as originally envisioned. Former fellows are now joining the ranks of Academy membership. Since 2009, 16 former fellows have been

elected as Academy members—just shy of one-quarter of the total number elected during that period.

We have increasingly viewed our mentoring task as an intergenerational effort.

During my time, we were extremely fortunate to add the Spencer Dissertation Fellowship Program to our portfolio. Spencer Dissertation Fellows were known nationally as a competitive, creative, and impressive group, and we could not have been happier to have these scholars—at the very beginning of their academic careers—join the Academy community. Our new Professional Development Committee was charged with overseeing the two selection and mentoring/retreat efforts and assuring their integration.

We also worked to increase the participation of Spencer/NAEd Postdoctoral Fellows—current and past—in our work across the board. First, there is a new avenue for the professional development of former fellows in the Dissertation Fellowship Program. Each application goes through an extensive review process by both the NAEd members as well as former fellows. We added former postdoctoral fellows to our selection committee and enlisted them in the external review of applications. In fact, they form the majority of the selection committee members and the application reviewers for the dissertation fellowship.

In addition to the selection of fellows, former postdocs are an essential part of the mentoring programs of new dissertation and postdoctoral fellows. Generally, former fellows comprise roughly one-half of all mentors and discussants at the fall and spring retreats, and they also compose nearly one-half of the retreat planning committee. We also tried to assure that each annual meeting panel incorporated current or former postdocs, providing fresh perspectives.

Now to the second aspect of building capacity. We are not only launching careers but also launching fields. I believe that our project on Adaptive Educational Technologies (AET) defined a new field of study: how the education research community can make use of the vast amounts of data that result from the use of these instructional technologies, and what concerns (such as privacy) must be kept at the forefront of everyone's thinking. Researchers had already been mining data from adaptive software for at least 15 years, but they had been mostly computer scientists and engineers, not education researchers. Our 2013 AET project served to bring education researchers to the table to study learning through this new window into student thinking.

Building Demand for Research

Making research more valued by policy and practitioner audiences remains a major challenge for the Academy. There has never, to my mind, been a more important time to ensure that education research is conducted with an eye to it being applicable to policy and practice—and to advance its use. As our politics become increasingly polarized, it can seem that evidence matters less and less. However, the Academy is one of the few groups that can confidently offer high-quality, unbiased, research-based interpretations and statements. We should continue to offer them, even as more and more voices with particular perspectives make policy debates noisier and more contentious. We must strive to be heard amid the clamor because what the Academy has to offer is unique and invaluable.

Our reports cannot go beyond their consensual conclusions—but we must find ways to push the policy implications of the research. Recently we have authored op-eds based on the NAEd reports, for example, one about the use of value-added methodologies in teacher evaluation and one calling attention to the necessity to consider concerns around big data use in education, including issues of security, access controls, and consent. Presentations at meetings and hearings offer additional avenues. We can make clear where the research consensus ends and the policy implications identified by the particular author or presenter begin, but stopping short of pushing the conversation limits the influence of the work.

We can also advance use by attending to the research on use. Historically much utilization research centered on translation, venues, and vectors—on how research is conveyed to policymakers. But more recently, the field has begun to think about how users make decisions and the role of research among other factors, drawing from studies such as the National Research Council's *Use of Science as Evidence in Public Policy* (National Research Council, 2012). As we move in this direction, I believe we must try to understand the qualities of research that affect use, such as the value of replication of findings in different contexts and the importance of syntheses that present the weight of the evidence (Fuhrman, 2000). Fortunately the latter is an Academy specialty, but we might also, as promoters of our field, try to influence university incentive systems so that synthesis and replication are appropriately valued.

Sustainability

Our mission of building capacity and demand for education research can only be addressed if we have sufficient resources. I close my brief reflections by urging more concern about the long-term health of the Academy, particularly financial sustainability. This is an area of work I regret we did not address sufficiently during my time as president. We

planned a development campaign, and then the fiscal crisis intervened. I believe we can return to this task with vigor in the future. We must continue to turn to foundations and agencies for project support, but we must also develop lines of work that can appeal to major donors for more significant support that can also contribute to the core functions of the Academy. If this were a “State of the Academy” speech, then I would say the Academy is strong, strong enough to take on the major challenges that remain.

References

- Feuer, M. J. Floden, R. E., Chudowsky, N., & Ahn, J. (2013). *Evaluation of teacher preparation programs: Purposes, methods, and policy options*. Washington, DC: National Academy of Education.
- Fuhrman, S. H. (2000). *AERA Division L Address, “Education Policy: What Role for Research.”*
- Gamoran, A., & Bruch, S. (2010). Alternative models for human capital development in education research. Retrieved from <http://www.spencer.org/sites/default/files/pdfs/GamoranReport.PDF>.
- Hedges, L., & Hanis, J. (2005). *Report on the evaluation of the NAE/Spencer Post Doctoral Fellowship*. Chicago, IL: The Spencer Foundation.
- National Research Council. (2012). *Using science as evidence in public policy*. K. Prewitt, T. A. Schwandt, & M. L. Straf (Eds.), Committee on the Use of Social Science Knowledge in Public Policy, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

The National Academy of Education: Personal Reflections

*Howard Gardner*¹

In the mid-1970s, I was a young researcher in developmental psychology and neuropsychology. I was living on soft money. With David Perkins, another young psychologically oriented researcher without a “real job,” I was co-leading Project Zero, a small research group focused on cognition in the arts. For historical reasons, Project Zero was housed in Longfellow Hall at the Harvard Graduate School of Education (HGSE)—though at the time, neither Perkins nor I nor our small research teams had much interest in education.

In the fall of 1974, Perkins and I, along with the noted philosopher Nelson Goodman, were invited to give a presentation to the National Academy of Education (then NAE, now NAEd), which was having one of its periodic meetings in Cambridge. (I do not think we had ever heard of the organization.) We used the occasion to talk about the research that we were doing in cognition and the arts. While I have no memory of the specific presentations, I am quite certain that they were scholarly in tone and were more appropriate, so to speak, for an arts and sciences faculty than for an education audience. Indeed, the other presenters were as prestigious a group as could be assembled: Seymour Martin Lipset, Alice M. Rivlin, Kenneth J. Arrow, and David Riesman. There were perhaps 20 to 30 attendees—mostly men, mostly grey-haired, mostly wearing the customary tweed suits and dresses. My guess is that our invitation came

¹ Howard Gardner is the Hobbs Professor of Cognition and Education at Harvard University. He was elected to the National Academy of Education in 1989.

from Israel Scheffler, a founding member of the NAE, and the primary reason was that Project Zero was housed at HGSE with the NAE. But I would not be surprised if this invitation had been vetted by Tom James, founding president of the Spencer Foundation, and/or by Pat Graham, who had recently moved to Harvard and was already a leader of the scholarly branch of the educational community.

Over the next decade or so, Project Zero and its leadership began a slow transition from being pure researchers “in the disciplines” to having a genuine interest in educational issues. In retrospect, the reasons for the movement are not difficult to specify. We were housed at a school of education; we were occasionally teaching courses there (without compensation, to be sure!); and, most important, educational issues and research were looming ever larger on the national radar screen. I was fortunate enough to have received grant support from the Spencer Foundation from the early 1970s, but that was at a time when founding President Tom James did not require that a researcher was an educationalist; and indeed, most Spencer funds at the time went to researchers in the social sciences or even in neuroscience. My other funding came from governmental agencies—National Institutes of Health, National Science Foundation, Veterans Health Administration, and soon, a new (and not long-lived) educational agency called NIE, the National Institute of Education. Within a decade, I ceased to secure funds from *any* governmental agency, relying increasingly on private foundations with a primary interest in education. I am quite sure that the availability of funds was also a prod for conducting research that had implications for education.

I was pleased and also surprised when I learned, in 1989, that I had been elected a member of the NAE. By that time, I was much more familiar with educational research and with educational researchers. I was fortunate enough to have remained a grantee of the Spencer Foundation, which, as other contributors to this collection will attest, has been the principal underwriter of the NAE over the decades. By nature, I am not particularly an “organization man.” Indeed I did not even join the American Educational Research Association (AERA) until my friend Elliot Eisner, by then the president of AERA, said, “Howard, you should be ashamed that you don’t belong to AERA.” But I sensed that the NAE was an organization that took scholarship seriously, and so I began to attend meetings and, upon request, to carry out little assignments.

Soon, I began my deepest involvement with the NAE. Accepting an invitation from Lee Shulman, I joined the committee that read and awarded Spencer postdoctoral fellowships. The committee, on which I served for 8 years and chaired for 4 years, was a lot of work. We had to read many applications, rate them, attend two meetings where we dis-

cussed the strongest applications, defend “our” choices, and ultimately award between 20 and 30 applicants.

Despite (or perhaps because of) the intensive work, I valued this experience. It gave me the opportunity to learn about what was being studied, and how it was being studied, by many of the most talented young researchers in the field. Equally valuable, it gave me the opportunity to work alongside leaders in the field, to understand their criteria for judgment, to debate with them on occasion, and sometimes to change my views as a result of those often lengthy discussions. It is not for me to say whether I succeeded in changing the minds of others.

After a 4-year apprenticeship on the committee, I became its chair. While it was not required (indeed, I was the requirer!), I read all of the grant applications, typically 200. It took up more than a few evenings and weekends! I believed that I could best participate in the decision process if I had at least a nodding familiarity with the range of applications, and only in that way could I spot trends, redundancies, and habits of various reviewers (both those on the committee and those in the NAEd’s far-flung network of non-committee reviewers). By the end of my term, I was prepared to turn over the assignment to others, but I cherish the experiences I had over nearly a decade.

In much of my working life, I have felt a bit of an outlier, or, as I sometimes quip, “I’m a member of the control group.” This characterization was true of my membership on the committee. Most committee members leaned toward awards to individuals with a clear interest in education, preferably those working in a school of education. In contrast, I leaned toward strong disciplinarians, whether or not they had a declared interest in education, and did not care whether or not they were housed in an “ed school.” Also, on the continuum with methodological rigor on one end, and a clever idea on the other, I tended toward the clever pole. Either I or someone else dubbed this “the wow factor.” In retrospect, over the 8 years, I became more sympathetic toward affiliations with education, and I think I nudged the committee members to pay more attention to the “wow” dimension.

I made one other contribution to the committee. Too often I felt that it was difficult to identify the nub of a proposal. Therefore, I convinced the committee that every applicant complete a short passage with wording something like this: “At present most individuals believe in XXXXX. As a result of the work proposed, most individuals will come to believe XXXXX.” Needless to say, this wording did not please everyone, and it was probably more appropriate for some topics and fields than for others. That said, I think that the sentence frame did push applicants to identify succinctly what was distinctive about the proposed work. And this was in the age before Twitter.

As part of my chairmanship of the committee, I became a vice president of the NAEd and attended leadership meetings regularly. My contributions were modest. Administration and governance are necessary and important, but they are not my thing. While it was occasionally mentioned by others, I had no interest in becoming the president. However, I remember engaging in late-night phone calls, when there were leadership crises—a sign that I cared about the organization's health.

One of the crises concerned the small staff of the organization. In a word, it was judged to be inadequate. I mentioned to Pat Graham that my daughter Kerith was interested in and good at administration of nonprofits. To my pleasure and delight, Kerith became the chief executive officer of the NAEd and, from all reports, served well during the administrations of Presidents Ellen Lagemann and Nel Noddings. I was especially delighted to watch Kerith—by nature a very reserved person—become focal and vocal in the lives of the predoctoral and postdoctoral fellows, as well as a young colleague on whom many academicians came to rely. Kerith went on to Yale School of Management, did a stint as a staff member at Teach for America, and is now an administrator in charge of faculty affairs at the Columbia School of Business.

When my term as chair of fellowships ended, I decided to reduce my involvement with the Academy. I take no pride in this decision, but I am not embarrassed by it either. We all go through changes in our interests and associations. What is important is that when called upon, one tries to do one's job well and to support those individuals who have leadership expertise. Indeed, after my years of active involvement in the NAEd, I joined the board of the Spencer Foundation and served for 10 years with other individuals, most of whom are the NAEd members. My membership on the Spencer board was greatly enhanced by my earlier service with the NAEd.

With both of these involvements concluded, one of my great regrets is that I no longer know who are the best young people in the field. Like the Spencer Foundation, the NAEd has always represented the gold standard in education—and at times when education is criticized and when support for educational research is scarce, the two organizations serve as reminders of educational scholarship at its best.

The National Academy of Education: The Center, the Margins, and the Future

*James Paul Gee*¹

We live in a high-risk and imperiled world. We face pressing problems from a myriad of complex systems spinning out of control. These are systems such as global warming, environmental degradation, massive inequality, disruptions from automation and new media, and a global speculative capitalism unmoored from productivity and devoted to short-term interests. Near constant global conflicts are exasperated by all these forces.

Our society has done too little to address these problems. Indeed, in a great many sectors, our society ignores and even disdains evidence. Some scholars have argued that the nature of our current institutions and even of the human mind itself (shaped as it is by the very different environment in which it evolved) are no longer well fit to face such complex problems. We need deep innovations and we need them soon.

Colleges and universities were the institutions that we once trusted and hoped would devote themselves to building new knowledge, solving hard problems, and honoring evidence over authority, ideology, or money. But our colleges and universities have been caught up in the same forces that are elsewhere endangering us all. Too often they are becoming businesses devoted to short-term gain rather than long-sighted ventures in the serve of knowledge and innovation. The number of tenure-track and tenured professors is dwindling. A large number of over-worked

¹ James Paul Gee is the Mary Lou Fulton Presidential Professor of Literacy Studies and Regents' Professor at Arizona State University. He was elected to the National Academy of Education in 2007.

and underpaid adjunct faculty are taking their place. Colleges spend the lion's share of their money on administrators and student recruitment and services, not on faculty or research. Grants seem to be becoming more important than the actual research they fund. Research in areas where grants are scarce is undervalued and unsupported. Furthermore, colleges and universities have, for the most part, stayed in their departmental and disciplinary silos and continue to teach in the same old ways—even as they move their lectures online—despite the hard problems demanding collective intelligence facing our world.

Yet, in some areas, the nature of academic disciplines, research, and expertise is changing. Some high-end research today tends to be organized in terms of cross-disciplinary teams dealing with a core challenge or “hard problem” that requires melding various methods, tools, theories, and technical languages. Single subdisciplinary, individual expertise is being superseded by more collaborative, cross-disciplinary, and collective-intelligence-based approaches to research. Furthermore, we have many calls today for academic institutions to put as much emphasis on impact as on publication, calls that cite the small impact much academic research has had in the past beyond building academic reputations (and “impact” here does not mean just “making money”).

In this complex setting it is necessary for organizations such as the National Academy of Education (NAEd) to push against the times and represent the importance of the traditional academic values of reasoning based on evidence in the service of the not always short-term values of knowledge production, debate, dialogue, discovery, and innovation. It is necessary, as well, for organizations such as the NAEd to push scholars in education to engage more and more in collaborative and cross-disciplinary research devoted to networking multiple disciplines, methods, academic languages, and smart tools in the service of solving some of the difficult problems we face today, especially as new digital media allow learning and teaching—and in some cases 21st-century skills—to spread well beyond schools, at least for some children.

In this piece, however, I want to focus on a third goal: the obligation to foster innovation when we need new blood to reinvigorate our various fields in a time of great change, complexity, and danger. However, as we face the difficult problems of our new world, issues of access still remain strong. People “at the margins”—whether this means members of groups that have typically not been given fair access; people at non-elite institutions; people poorly mentored at elite institutions; or people engaging in work that lies at the margins of current paradigms—still often lack the social networks to gain access to opportunities to grow, contribute, and flourish. Yet, in times of change—and at a time when much work in our many areas of education has run its course, become repetitive, or has

failed to truly impact and sometimes even really address the problems we face—innovation and paradigm changes may well come from the margins not just the center.

We in education have worked diligently for decades on the problems of access, inequality, various “gaps” between different groups, and the stolid inertia of our schools, classrooms, and educational institutions. However, these problems persist today without demonstrated improvement from the 1960s. Indeed, inequality is at an all-time high. Segregation is worse than it was in the 1960s. The literacy gap by class has passed the gap by race, not because we improved the race gap all that much but because we made the class gap worse (in part by ignoring it and too often confounding race and class).

It is time now to stop business as usual. It is time now for real efforts to grow new theories, methods, and approaches in a timely fashion. It does not seem, however, that these new ideas are likely to come from the center alone, given that most of us NAEed members got to center from a past far different from the present.

Education has a number of prestigious grants, fellowships, and prizes for graduate students and young scholars. These tend to be dominated by people either from elite institutions or people working at the very center of their fields (usually both), often dedicated to incremental contributions to now fairly well-established bodies of research. There is nothing wrong with this. It is how science develops at times of relative stasis. Unfortunately, we are not living in such times. It is time now to marshal real attempts to foster innovation, to take risks, and to cultivate the margins, in addition to business as usual. Many young scholars take the word “margins” to be an insult, but the task here is not to make students and scholars at the margins “mainstream,” the task is to have the courage sometimes to marginalize mainstream work so that innovation can happen and the margins can shift.

We need now, sometimes, to use selection processes that seek out risk, innovation, and new talent in order to find the new currents that show promise to refresh and transform the mainstream research and interventions in education. We need to create spaces where risk taking and failure can flourish and young scholars can get ready for a “prime time” that will be constituted of new shows and not repeats and minor riffs of old ones.

This is not affirmative action. It is not about bringing disadvantaged people into the fold so that they can be like us. It is about realizing that the future will not be like the past and so now our best hope may well reside somewhere at the margins. It is about action on the mainstream’s part to acknowledge not only their very real merits but also their failures and the massive changes in our world. This is not about political correctness, it is about evidence and respect for a real and dying world we all live in. It

is about what all sports teams past their prime do: look to the bench. It is about not believing that the elite players we have drafted thanks to their outstanding “stats” are not necessarily always better than a walk-on who changes our game for the better.

Mentoring would play in developing new talent, but not mentoring that mentors young scholars just to be more mainstream. It is not about treating young talent from the margins as “at risk,” but, rather, as forces that can help us all change some paradigms. Times of great change are, for many older academics—or should be—times not only of risk and fear of falling, but also of real excitement in the search for new ideas, new promises, new failures, and a legacy in which our labors as now older academics will become part of the ongoing and changing histories of our fields and not the ends of them.

The reason to do this—to actively seek promise at the margins and to actively question the current depth and innovation at the center—is not morality (though that would be good reason enough). The reason is survival and the fulfillment of our intellectual obligations as professors to ensure that knowledge—and not lies—grow and flourish even when this means distrusting our own long-honed judgments about what is at the center and what is at the margin. I am not saying that our problem is that the mainstream in many areas of Education are Dead White Males (nor am I denying it). I know a number of mainstream women and a number of mainstream African-Americans (and some people who are both). I am just saying that the mainstream is in some important areas and in some important ways in need of refreshment, as is our world. I hope and believe that the NAEd can be a force for holding sacred the crucial accomplishments of our past, even when they are not trendy or financially profitable; a force for pushing forward new forms of collaboration, methods, and tools to create collective intelligence to speak to hard problems in an impactful way; and a force for scouring the margins and taking risks in the service of innovation so we can revitalize education at a very exciting, but demanding, time.

Presidency of the National Academy of Education: 1985–1989

*Patricia Albjerg Graham*¹

My presidency of the National Academy of Education (NAEd) began in 1985. I received a title and a single shoe box that contained the Academy's records. So it began.

The NAEd was then 20 years old, and in its earliest years had engaged many distinguished scholars, most outside of schools of education, in conferences and publications. Perhaps best known was its *Impact of Research on Education*, edited by Patrick Suppes and published by the Academy in 1978, known to the membership colloquially as "the big blue Suppes." Though money was always tight, funding did support a few tiny fellowships, commissioned papers, and convened conferences. As President Robert Glaser, a legendary psychologist and head of the Learning Research and Development Center at the University of Pittsburgh, and his highly competent administrative assistant, Michelle von Koch, contemplated the end of their service, the Academy's financial future looked bleak. Now 2 years after publication of *A Nation at Risk*, Foundations, which had supported educational research in the past, seemed less interested in conferences or 679-page volumes on educational research and more focused on issues of educational practice. The NAEd's funds were nearly gone.

However, Lawrence Cremin, a principal founder of the Academy in

¹ Patricia Albjerg Graham is the Charles Warren Professor of the History of American Education, Emerita at Harvard University. She was elected to the National Academy of Education in 1979.

1965 and a subsequent president, did not want to see the Academy die. Twenty years later as outgoing president of Teachers College, Columbia University and incoming president of the Spencer Foundation (and a long time member of its Board), Cremin was in a unique position to assist the Academy. He knew well most of its members, and thus could approach them to consider the presidency of the Academy. One so approached could then reply that acceptance of this position was dependent on support by the Spencer Foundation. Cremin's reply, of course, was that a proposal must be submitted worthy of Spencer funding and approved by the Spencer Board of Trustees.

Thus, my phone rang one spring day in 1985. Cremin, who had been the principal advisor for my PhD, urged me to accept the nomination for presidency of the Academy. I was reluctant, given other commitments I had at the time. Furthermore, I knew that it would be impossible to revitalize the Academy without administrative help and that assistance must be tied to legitimate work of the Academy. This required serious funding.

At the time I was Dean of the Graduate School of Education at Harvard University, and it was achingly evident to me how difficult it was for either women or men to complete the scholarly work needed for tenure during their "apprentice" years. A fellowship that would permit them to work full time on their research would enable many to complete the scholarly work that would make them strong candidates for tenure but even more importantly develop them as lifelong scholars and investigators of educational dilemmas. The loss of research interest and expertise that occurs during beginning faculty years in institutions with heavy teaching loads is immense. That loss is also wasteful of the wisdom that such faculty already possess and could augment with support for further investigations. The Academy and Spencer had previously engaged in small numbers (five per year) of fellows with very limited stipends, not enough to quit teaching to work on a manuscript. What I proposed to Cremin was a much larger cohort of 25 or so fellows fully funded for two semesters, drawing both on individuals prepared both in education schools and in arts and sciences or other professional fields. Because I was also on the board of the Spencer Foundation, I knew that I must recuse myself from any Spencer action and that the Foundation had high standards for projects it would fund.

During the summer of 1985 Michael Fultz, then a doctoral student at Harvard and working part time as an assistant in the dean's office, sat with me on my front porch in Cambridge and together we developed a formal proposal to Spencer for a two-semester, full-time fellowship program for individuals within 5 years of their doctorate. We established meetings of the fellows concurrent with the NAEd so that the young (the fellows) and the "walruses" (as Academy members referred to themselves) could become acquainted with each other and each other's research. Spencer

agreed to fund the proposal and has been doing so with modifications and improvements ever since. Key to our success was hiring Gail Keeley as our executive officer, who managed the Academy magnificently.

During the 1985–1986 academic year, under the wise and compassionate leadership of Lee Shulman the Academy announced and administered the postdoctoral fellowship competition. Twenty-nine fellows were selected, several of whom are now NAEd members, and they are an illustrious group, as are their successors. This was the center piece of the Academy's activities during my presidency from 1985 to 1989, when I was succeeded by Shulman and the Academy moved to Stanford University, where Shulman was a professor.

The postdoctoral fellowship program raised the NAEd's visibility and created other opportunities for projects of interest to the Academy and to supportive funders. The first of these was a review of the National Assessment of Educational Progress (NAEP) conducted by Governor Lamar Alexander and Tom James, former dean of the Stanford School of Education and former president of the Spencer Foundation. The commentary was prepared by Anthony Bryk, then of the Harvard Graduate School of Education and at the University of Chicago when his analysis was completed. The success of this enterprise led to a number of solicitations from the U.S. Department of Education for Academy involvement in priority setting for its Office of Educational Research and Improvement, most of which the Academy declined.

The Academy also participated in a number of exchanges with educators in the Soviet Union at a time of "Glasnost" in that country when educational and political practices were in flux. Formal delegations of Academy members visited the Soviet Union in 1986 and 1988, and the Academy received several Soviet delegations. In addition, a number of smaller meetings between Academy members and Soviets took place during this period. Subsequently the new Russian government showed less interest in these exchanges.

The landscape of educational research was beginning to change. The emergence of independent organizations soliciting funds to advise on educational matters, including research, challenged the NAEd's supremacy in giving educational guidance. Nonetheless, from 1985 to 1989 the Academy raised \$5.3 million from Spencer, Carnegie Corporation, Exxon, Mellon, Rockefeller, Ford, Matsushita, Hewlett, and MacArthur foundations largely for analytical papers on educational research priorities or analyses of complex educational problems. Funding for educational research was also changing. The U.S. Department of Education reported to the Academy in 1988 that its funding for educational research had taken a 70 percent cut. In that context Lee Shulman, the next president, and his successors continued the postdoctoral fellowships but also found new sources to continue the Academy's leadership on educational issues.

Membership in the National Academy of Education: An Ongoing Dialogue

*Susan Moore Johnson*¹

One dreary winter day in 2000 a splendid bouquet of flowers arrived at my office. Because I had nothing particular to celebrate, I thought at first that it must be meant for someone else. An enclosed card from a former doctoral advisee offered warm congratulations, but no further explanation. So I called to say thank you and to ask “why me?” It turned out that her partner, who had attended a National Academy of Education (NAEd) annual meeting with his mentor and collaborator, happened to hear that I had been elected to membership. She assumed that this was not news to me. It was, and so I kept it to myself until the official notice arrived about a month later.

I knew very little about the inner workings of the NAEd at the time, although I had already benefited from its work, having received an NAEd/Spencer Postdoctoral Fellowship in 1987—the second year of that competition. Pat Graham, former NAEd president and then Spencer Foundation president, had seen potential in the two organizations collaborating to create new opportunities and support for early-career scholars. Because I was not a traditional researcher, applying for a postdoc seemed like a long shot at that time. I had spent nearly a decade as a high school teacher and administrator, which meant my research experience was limited and my curriculum vitae short. In addition, I used qualitative methods to study the intersection of teachers, school organizations, and policy, an approach

¹ Susan Moore Johnson is the Jerome T. Murphy Research Professor of Education at Harvard University. She was elected to the National Academy of Education in 2000.

and focus that were uncommon at the time. Although I thought my work was important, I felt very much on my own. Receiving that postdoc not only fueled my career, but also my confidence in pursuing an important line of research using methods that were hardly mainstream.

Although I was deeply indebted to both Spencer and the NAEd for the opportunities opened by the postdoc, I had never considered what it might mean to be part of the NAEd until those flowers arrived. The organization and those who guided it had changed the course of my professional career, and I decided I could pay that gift forward. However, my interest in being an active participant in the NAEd was not entirely selfless. Reviewing the list of the NAEd members, I could see that this revered organization had considerable strength in traditional academic disciplines, such as psychology, economics, history, and philosophy. However, there were relatively few members who focused on problems of practice within schools and who saw merit in addressing those problems from a multidisciplinary perspective. I hoped that over time the disciplinary boundaries within the NAEd would become more permeable as an increasingly diverse group of scholars collaborated inside and outside the organization to address some of education's most pressing and important challenges.

When I was elected to the NAEd board in 2003, I discovered that others were searching for ways to expand the membership as well. Under Lorrie Shepard's leadership, a subgroup of board members convened to review the fields of research currently represented in the membership. We then urged individual members to identify and nominate deserving scholars who might be missing from their current field. This informal process led to the election of a strong group of new members, whose interests, inclinations, and research contributions enriched the array of the NAEd talent. Importantly, it also brought in new members who were committed to working for the organization.

Under Susan Fuhrman's presidency, the board took steps to create what we informally called "cradle-to-grave" opportunities for scholars. This involved systematically and continuously linking individuals who had received NAEd/Spencer Dissertation Fellowships and NAEd/Spencer Postdoctoral Fellowships with current members. There was, of course, no certainty that receiving a fellowship would lead eventually to the NAEd membership and no barrier to electing members who had never received a fellowship. What we envisioned was an ever-increasing pool of outstanding scholars and leaders of practice whose favorable experiences with the NAEd would encourage them to participate in the organization's work as mentors, reviewers, participants on panels, and collaborators on research projects. To launch this work, the board established a standing professional development committee. As its first chair,

Marilyn Cochran-Smith very ably led this committee's work for the first 4 years, giving life and meaning to the board's idea.

However, just as it is the tenured faculty who determine the shape and direction of a university department, it is the elected members of the NAEd who define its future. At an annual meeting in 2011, several members raised concerns about what they perceived to be a lack of diversity in the membership and an opaque, potentially discriminatory election process. The NAEd bylaws authorize the board to appoint vetting committees for each nomination and review their reports, decide which candidates to include on the ballot, and make final election decisions about membership after the vote. Those who spoke at that meeting suggested that the time-consuming task of preparing a nomination might be quickly undone by a secret vote of a biased board. In quick response to these concerns, Susan Fuhrman and the board formed a membership committee to review the current process for nominating and electing members and to make recommendations for improving it. As the NAEd's secretary-treasurer, I chaired that committee, which included Hy Bass, Marilyn Cochran-Smith, Margaret Eisenhart, Larry Hedges, Gloria Ladson-Billings, and Mike Smith.

Early in our committee meetings, we sought more information from the NAEd staff about the profile of its membership. Greg White and Philip Perrin prepared a detailed analysis of trends over time. They found that overall the membership was aging, largely because of the increasing average age of members at the time of their election. Furthermore, there continued to be a notable lack of racial and ethnic diversity, and gender imbalance persisted. Members also were concentrated in a relatively small number of geographical regions and universities. After presenting this information to members at a 2012 meeting, we began to propose changes in the election process that we thought would maintain high standards while also expanding the fields of knowledge, approaches to inquiry, demographic diversity and extent of members' engagement with policy and practice. These included providing reassurance to the members that, despite what the bylaws allowed, the board would not be arbitrary in deciding who would be included on the ballot or how votes would be weighed. We actively encouraged members to consider nominating younger scholars, candidates of color, and those from a broader range of universities. We revised instructions for members of vetting committees, asking them to judge whether a candidate deserved to be included on the ballot, not whether they thought that an individual should be elected to membership. Overall, we wanted the board to be advised about each nomination by experts in the field, while not disenfranchising members. Our changes satisfied some members, but dissatisfied others, who worried that they might lead to a ballot of compromised quality. These altera-

tions in the nomination, vetting, and voting process, along with a tighter timetable that conformed better to the academic calendar, yielded a larger and more diverse group of members in the next election.

The committee also considered longer-term issues. For example, the NAEd has long included in its membership a small number of nationally recognized leaders in policy and practice. Recently, however, those numbers have declined, despite efforts by the committee to encourage such nominations. Without addressing this issue, the NAEd is likely to be less influential in the worlds of policy and practice that it hopes to influence. However, our committee realized that it is difficult to elect such individuals when many—if not most—members rely on conventional, scholarly criteria for electing members. Creating separate categories for scholars and leaders in policy and practice might address part of the problem, but also have the unintended consequence of creating a two-tiered system that left leaders of policy and practice with less recognition and influence. The committee did not find a good solution, but did highlight the continuing importance of this issue.

The membership committee also discussed the potential effect of raising the cap on membership, an issue that is currently under discussion. It seemed possible that, with a larger NAEd, key aspects of the nomination, vetting, and election process could be better managed by subgroups of members who are experts in a field. Hy Bass provided valuable information about the election process in the much larger National Academy of Sciences. We also learned that, years earlier, the NAEd had been divided into discipline-based “sections,” which were subsequently eliminated because, as one member said, they “caused internecine battles” between historians and philosophers. However, because those were unlikely to recur today, it seemed likely that a growing membership eventually would require us to devise a more decentralized nomination and vetting process. Finding a good way to do so, while also promoting interdisciplinary exchange, continues to be a challenge. As a first step, we asked all members to select one or two fields from a list that included both disciplinary and interdisciplinary options. In the short run, these lists can guide the board and secretary as they appoint vetting committees; subsequently they might serve as the basis for a revised nomination and election process. Currently, under Michael Feuer’s presidency, Catherine Snow is chairing a new membership committee as we work our way to establishing an election process for that NAEd that is inclusive, rigorous, and fair.

The growing size and changing composition of the NAEd, along with its increasingly active role in developing opportunities for outstanding researchers at all stages of their career, make it more likely that the organization can meet its goal of “advancing high-quality education research

in policy formation and practice.” To do so, however, will require ongoing efforts to refine the election process and discover new ways to engage the full array of members, current fellows, and former fellows, in joint inquiry, exchange, deliberation, and action on behalf of improving education for all students.

Recollections of My Presidency of the National Academy of Education, 1993–1997

*Carl Kaestle*¹

Just before I began my term as president of the National Academy of Education (NAEd), the Academy received an invitation from our colleague Torsten Husén to hold a meeting of the NAEd in Stockholm. This generated a flurry of activity. Very helpful support came from the Spencer Foundation, whose officers subsidized our desire to bring some of our young fellows along. Support also came from our colleague Professor Husén, who provided many necessities for our stay in Stockholm. We had a robust conference of papers and discussions. I recall the well-attended session with Jim Comer, who considered a question from the audience: could a white, middle-class reformer achieve the kind of grassroots parental engagement that characterized his reforms in New Haven, especially if he had not grown up (as had Comer) in a tough African American community? We also had many ceremonial occasions, presenting gifts to the mayor of Stockholm, the president of the Royal Swedish Academy of Sciences, and others. We used the impressive meeting room of the Royal Academy for several sessions. This joint meeting was rich in exchanges; we were received with exceptional pomp and lovely occasions.

A second priority early in my term was to find a permanent home for the business office of the NAEd. Under the existing system the office moved every 4 years to the president's institution. This made it impossible to retain staff members from one presidency to the next. A half-

¹ Carl Kaestle is University Professor of Education, History and Public Policy, Emeritus, at Brown University. He was elected to the National Academy of Education in 1981.

measure had been taken just before I began my term as president. During a meeting in Chicago, attended by board members and past officers, we discussed whether we should continue reconstituting and moving the staff every time the president changed, or locate the NAEd in one permanent location. The most avidly discussed issue was whether we should establish a Washington headquarters. There were those who argued that we were an academic organization, not a policy shop, and therefore we should be based at a university. In Washington, some asked, who would be the community of colleagues for our staff? Would it not risk skewing our mission toward policy research and promotion? The National Research Council (NRC), as I recall, was not mentioned. The wind was not yet blowing in that direction for our ship.

Shortly before I took office, a compromise was reached. We should find a university willing to afford us space for a permanent office (a request that university officials were increasingly reluctant to welcome). Lee Shulman, my predecessor, suggested that I give it a trial run with a letter to some likely major universities with substantial numbers of the NAEd members on their faculty. We received two offers, and the board decided that Stanford would be the best. Thus we resided at Stanford throughout my term. Being in Chicago, I was the first president who was not co-located with the office.

Further discussion continued after I left the president's position, and eventually a proposal emerged, with support from Michael Feuer, a rising star at the National Research Council, and a nice collection of National Academy of Sciences (NAS) members who favored the idea. My successor, Ellen Lagemann, I believe, represented the NAEd in these discussions and in formulating the proposed liaison. (Ellen succeeded my actual successor, Ann Brown, who died tragically some months after she began as president of the NAEd.) The cooperative arrangement provided us with space and collegueship at the NRC's headquarters. However, the NAEd is not governed nor funded by the NAS and not designated as one of the academies in the National Academies. The advent of Bruce Alberts as president of the NAS also sweetened the deal. Alberts has strong interests and experience in science education and made our transition there all the more welcoming. I believe that this arrangement has indeed worked out very well.

Another priority during my presidency was to expand the membership of the NAEd and diversify it with regard to neglected fields, and new fields, as well as with regard to women candidates and scholars of color. My being at Chicago was serendipitous for this task because that is where Charles Bidwell was. Charles was an officer of the NAEd, a prominent sociologist of education, and a leader in the University of Chicago's renowned Department of Sociology. He was a good friend and a consum-

mate academic bureaucrat. We would meet for lunch at the Quadrangle Club (the faculty club) to talk shop about the NAEd. I did not think the food was so great, so I called it the "Triangle" Club. However many legs it had, that was where Charles and I laid the groundwork for a change of structure for the NAEd.

We thought that if we were to expand the membership, and fill gaps in interdisciplinary fields, then the disciplinary sections into which our members were organized would have to be reformed. Nominations emanated only from the sections. Some sections submitted a lot of nominations, others hardly any. It was not a system that would easily realize our collective goal to fill in disciplinary gaps, find and engage more scholars, and include more diverse scholars. Soon Charles and others had carried the day. The termination of the disciplinary sections was well-placed in Charles' hands. Some member quipped, "Who better than Charles Bidwell to discontinue the sections; it's like Nixon going to China." And it succeeded.

We then proposed and received support to raise the membership ceiling. We were, as I recall, still at the original 60 (15 for each section), and I think we raised it to 100. This was the beginning of a process; subsequently it was raised several times. Today, the NAEd is much larger, more diverse, and more able to carry on several projects at one time.

Two other projects related to the NAEd's capacity for policy-relevant research. The first was already in progress before I took office. It was a committee to study and report on the effects—intended and unintended—of the "state NAEP." The National Assessment of Educational Progress was conducting a trial run, in which the test was administered to statistically appropriate samples of students in all the states and the scores were reported at the state level, not just the national level. Our committee included many testing experts, with staff support provided by the American Institutes for Research in Washington, led by George Bohrnstedt. The co-chairs of the advisory committee for the evaluation of the "state-NAEP" were Robert Glaser and Robert Lind, the formidable "Two Bobs" of testing research. Many theory and policy issues were involved; for example, some testing experts thought that if NAEP was to stand as the "Nation's report card," one could not start disaggregating to states, or districts, or (ultimately) schools or individual students. If NAEP started to have consequences at lower levels, then that would contaminate its use as a national index of learning. The NAEd's evaluation (in four volumes) received some criticism, and some of it was political. My reaction was that it was sound, thorough, and constructive.

Talking it over with the "Bobs," I thought it would be appropriate for me to prepare for this project by attending the working sessions, which were quite interesting. Despite my murkiness about the fine points, this

stood me in good stead for the few times that I spoke on behalf of the NAEd about the work of the committee (once at the National Press Club). In addition, it helped me learn more about the field of assessment, one of the benefits of belonging to an interdisciplinary group.

The second research-based NAEd project that engaged me in my 4 years as president concerned the Clinton administration's proposal to create a National Education Standards and Improvement Council (NESIC). The Clinton education team proposed a relationship between the states and the federal government to initiate standards-based-reform. Its Goals 2000 legislation proposed that states would establish state standards and align assessments with them. On a voluntary basis, they could submit the standards to a new board. After the bill passed, some people in the education policy business, in Congress, foundations, and the research community thought that many members of Congress and others selected for NESIC membership, might not know very much about standards-based education.

Representatives from three foundations (Spencer, Pew, and MacArthur) talked to me and offered a small grant for us to prepare an introductory publication on the nature and subtleties of standards, with a view for a subsequent larger project for the NAEd about current school reform concepts. I appointed a committee of about eight NAEd people well-versed in policy studies to consider the matter. From those discussion emerged a monograph with Milbrey McLaughlin and Lorrie Shepard as co-authors and with Jennifer O'Day as staff writer. It was titled *Improving Education Through Standards-Based Education* (1995), an expert, concise policy book about the subject, with emphasis on concepts, challenges, and cautions. Its intended audience, however (potential members of NESIC and others interested in its work) disappeared when Republicans in Congress refused to fund the NESIC, essentially overturning this aspect of Goals 2000. I hasten to add, however, that many education reformers and researchers read this gem, and it is still very much worth reading.

Following the publication of the report, our committee reconvened a few times to consider the invitation to undertake a longer-term, larger work on education reform. We had some pretty good ideas; I presented some of them at the Pew Trust at a small gathering, with some positive feedback. In the end, however, this project faltered because every NAEd member whom I thought had the scope and experience to co-chair such a study was too busy with project leadership elsewhere. This underscored the fact that our membership was just beginning to grow, and our group of members with deep knowledge about federal and state reform policies was limited and prodigiously busy. Despite our limitations, then, the NAEd did some important work in collaborative policy studies during this seed-time of standards-based reform.

These policy studies were carried on by a small group of our members. In the meantime, the life of the NAEd in other pursuits was energetic. We began the process of electing a wider membership, including those in basic, theoretical, historical, and other approaches to education. We continued the time-consuming and enjoyable work required to operate and improve our postdoctoral fellowships supported by the Spencer Foundation. (The dissertation fellowships were already launched but not administered by the NAEd at this time.) Panels at our meetings covered a range of theory, practice, methods, and policies.

In closing, I am happy, in this my 35th year in the NAEd, to see how richly endowed we are with a larger, more diverse group of scholars across the spectrum of disciplines, methods, and concerns, with a wonderful permanent home and a professional staff. Happy Birthday, NAEd!

The National Academy of Education as a Bold Source for Big Ideas

*Ellen Condliffe Lagemann*¹

The National Academy of Education has an interesting history. According to Larry Cremin, he and Israel Scheffler “dreamed up” the idea of an Academy during the summer of 1964, when he was teaching at the Harvard Graduate School of Education and was housed in an office near Scheffler’s in Longfellow Hall. Apparently, the two got together for lunch every day and talked about big ideas as they munched bologna sandwiches—a somewhat unlikely diet for two Jewish boys from New York. At the time this was reported to have occurred, Larry was developing the critique of the history of American education that he was soon to publish as *The Wonderful World of Ellwood Patterson Cubberley* (Cremin, 1965). Agreeing with his friend Bernard Bailyn, the distinguished Harvard colonial historian, that the history of education was anachronistic in its exclusive concentration on school history, Larry was eager to move the field from a narrow professional focus to a broader, cultural perspective. That ambition stood behind much of his scholarly work, including his three-volume *American Education* (Cremin, 1970, 1980, 1988), and also motivated many of his professional activities, including his role in designing the Academy.

Although the Academy was founded “on the initiative” of J. B. Conant, former president of Harvard, and John W. Gardner, president of the Carnegie Corporation of New York, it was Larry who put them

¹ Ellen Condliffe Lagemann is the Levy Institute Research Professor at Bard College. She was elected to the National Academy of Education in 1990.

up to launching the organization. It was also Larry who took the lead in identifying initial members, writing the initial constitution (which was modeled on that of the National Academy of Sciences), and planning early meetings. It was not by coincidence that the Academy began its life with a heavy representation of scholars from the humanities as well as the social and behavioral sciences; nor was it by coincidence that the original "sections" were defined by those disciplines rather than according to the sub-fields of education. The Academy was established to embody and advance the kind of discipline-based scholarship in education for which Larry advocated throughout his career.

In establishing the Academy with a preponderance of discipline-based scholars, many from arts and science departments, Larry hoped to winnow the field. He wanted to give prominence to those scholars whose work met standards of disciplinary excellence and to gain standing for the Academy as an arbiter of what counted as high quality. Clearly, there was a considerable dose of elitism in these ambitions. They were of a piece with "the best and the brightest" mentality of the early 1960s. But there was also a wish to advance a distinctive approach to the study of education, one in which education was seen as the means through which large national purposes were defined and advanced. Larry was not interested in policy studies narrowly constructed around questions of feasibility and strategy, or "what works." He thought scholarship in education was most significant when it focused on matters related to what he liked to call the American *paideia*—"a vision of life itself as deliberate cultural and ethical aspiration" (Cremin, 1979, p. 19). The Academy, with its stellar roster of scholars, was intended, in his view, to support that kind of scholarship.

Policy study defined in this broad way is not much in vogue these days. Policy study is now more likely to focus on the success or failure of current school reform activities than on larger questions of national purpose and direction. This may reflect the current popularity of evaluation as an approach to policy analysis as well as widespread discounting of insights derived from philosophy and the humanities. Whatever the causes, as it begins its second 50 years, I hope that the Academy can play a part in fostering broad discussion and debate about questions concerning the role of education in addressing the many challenges facing the United States.

Inequality has increased dramatically in recent decades, as have tensions between minority communities and representatives of government authority, especially the police. The U.S. Congress and many state legislatures are paralyzed. The nation's physical infrastructure has been dangerously neglected, as the price of waging wars and operating the largest prison system in the world has risen. Problems such as these are not solely educational problems, but they cannot be adequately resolved

without education. Civic education must be redesigned inside schools and colleges as well as outside. Ways must be found to teach tolerance much more powerfully than has ever been done before. Learning about the science relevant to everyday life, whether pertaining to global epidemics or the importance of composting, must become a lifelong task, involving the popular media as well as more traditional “teachers.” Figuring out how to consider and move toward action on questions such as these probably sounds far-fetched. But devoting some of our collective time and energy to thinking and talking about education in relation to big problems could offer the Academy a niche that does not place it in competition with organizations better equipped to engage in policy analysis of a more traditional kind. It would align nicely with our history—and, perhaps most important, it could be interesting, fun, and perhaps even useful.

There was a fair amount of chutzpah involved in establishing the Academy—in gathering a small group of scholars, who designated themselves as exemplary, and took it upon themselves to review and evaluate work across the broad field of education scholarship. Following in the spirit with which the Academy was founded, I hope that the Academy will be a bold source of big ideas during its second half-century.

References

- Cremin, L. A. (1965). *The wonderful world of Ellwood Patterson Cubberley*. New York: Teachers College Press.
- Cremin, L. A. (1970). *American education: The colonial experience, 1607–1783*. New York: Harper & Row.
- Cremin, L. A. (1979). *Traditions of American education*. New York: Basic Books.
- Cremin, L. A. (1980). *American education: The national experience, 1783–1876*. New York: Harper & Row.
- Cremin, L. A. (1988). *American education: The metropolitan experience, 1876–1980*. New York: Harper & Row.

Thoughts on the National Academy of Education 10 Years Later

*Nel Noddings*¹

I served as president of the National Academy of Education (NAEd) from 2001 to 2005. It was a busy time. We hired two new executive directors during that time (both terrific people), and in 2005 we moved our office from New York University to the National Academies building in Washington, DC. We also launched the Annual Fund for the Academy's Future and raised a little more than \$20,000 in our first attempt. It was worth trying, but we continued to need the help of generous foundations, and we have been grateful especially to the Spencer Foundation for its support of our highly successful Fellowship Program. It is deeply satisfying to follow the careers of outstanding educators who started out as NAEd/Spencer Fellows.

Among other projects undertaken in those years, the NAEd co-sponsored with Kappa Delta Pi and the National Society for the Study of Education (NSSE) a series of forums on the No Child Left Behind Act of 2001 (NCLB). These were well-attended, exciting meetings. Rereading the President's message I sent in Academy Notes (November 2004), I am reminded of how concerned we were about NCLB and its likely effects. I wrote,

Although no one has spoken in favor of NCLB at the forums I've participated in, I've heard some favorable comments from a few superintendents of small city schools. They have told me that NCLB has given them

¹ Nel Noddings is the Lee Jacks Professor of Education Emerita at Stanford University. She was elected to the National Academy of Education in 1996.

the authority to shift resources to poor and minority students—that is, to promote practices they have always wanted to try but could not because of the demands of wealthier constituents.

I wonder how those optimistic educators feel now. Unfortunately, the hope expressed by these administrators has largely faded in the years since, and we still desperately need to do something not only about funding gaps but also about the growing communication gap across social classes in our society.

We were also worried (rightly) about the growing emphasis on national standards, testing and the mandated use of specific methods—even scripted lessons. In one of my messages, I wrote against the crazed effort to find one best lesson form, one best way to teach everything to everybody. Instead, I suggested that we advocate widespread teacher-tinkering—a commitment to the responsible use of a variety of teaching methods. I specifically mentioned the work of Hugh Burkhardt and Alan Schoenfeld on the “engineering” approach in math education, of Jerome Bruner on discovery, and of Lee Cronbach’s sage advice to cease the search for grand generalizations on lessons. In the NAEd’s newsletter, *Academy Notes* (May 2003), I noted that Cronbach advised us to seek and try out limited generalizations:

methods that seem reliable with topics like this one, with kids like these, in contexts like this.... The engineering method itself seems important. What is the big idea here? How does the exercise or lesson we are considering fit with the big idea? Is there a method that is especially compatible with this lesson? Are there collateral benefits to be gained? After trying and idea out, evaluate our experience and the results, tinker some more, try again.

Today, I still endorse this approach, but I would add that there are some days when the prepared lesson should be discarded entirely and replaced by exploration of a question raised by a student. (Remember David Hawkins’ “How to Plan for Spontaneity”?) That question might engage fresh intellectual interest in the subject under study or it might help to connect the subject to other subjects in the curriculum, or it might induce discussion of what it means to live a moral life, what it means to be a citizen, to be a friend, to be a parent. I still believe there is more to teaching than instruction, certainly more than getting high test scores.

If We Know So Much from Research on Learning, Why Are Educational Reforms Not Successful?

Lorrie A. Shepard¹

In this essay, I return to the question that framed my presidential address to the National Academy of Education (NAEd) in 2009. Given a deep research base on learning, why are educational reforms not successful? At that time, a new President Obama was beginning his first term in office with great hope, and during the preceding election year the NAEd had provided to both political parties a series of white papers summarizing research relevant to key education issues. One of those white papers, *Standards, Assessments, and Accountability* (Shepard, Hannaway, & Baker, 2009), was the focus of my address titled “Curricular Incoherence: The Story of Educational Reforms Undone.”

Back in 2009, my intention had been to explain the bait-and-switch errors that arose in previous reforms when similarly named policies were substituted for research-inspired ideals. My hope was to forestall the problems that arise from superficial understandings. Now with 6 years elapsed, we see disappointingly that cautions—issued by many—were not heeded, and ill effects from top-down accountability mandates continue to pervade the education landscape. To understand why these patterns continue and what might be done about it, I repeat again the old history of standards-based reforms in the 1990s, attending in particular to the connections that reformers drew then to advances in cognitive

¹ Lorrie A. Shepard is the Dean and Distinguished Professor of the School of Education, University of Colorado Boulder. She was elected to the National Academy of Education in 1992.

science and research on learning. I then explain how reforms have been undone by direct attacks, but more often by competing visions; or they are subverted by superficial understandings and lack of support. In the last section of the essay, I repeat the most important of the NAEd white paper recommendations regarding standards and assessments. In the case of assessment reforms, I try to explain why short-cut versions of imagined reforms have again fallen short of what might have been possible; I also consider which if any of the best ideas in that paper might still be worth pursuing. I conclude with a plea to roll back accountability mandates, which have only exacerbated inequities, and to invest instead in smaller-scale reforms designed to support teaching and learning.

The term “standards-based reforms” was coined in the 1990s to signify the central role that “world-class standards” were expected to play both in raising expectations and in crafting coherent systems to meet these higher standards. A spate of policy reports condemned the existing, de facto basic skills curriculum driven by standardized tests and textbooks and called for the creation of challenging standards aimed at higher-order thinking and problem-solving abilities. As part of an aligned system, new forms of authentic assessments requiring more fulsome enactments of ambitious learning goals were expected to leverage reform efforts instead of misdirecting teaching and learning as previous tests had done.

Leading education researchers were deeply involved alongside policymakers in arguing for standards and accompanying reforms. Mathematics educators led other disciplines in developing curriculum and evaluation standards that sought to change the character of what mathematics was thought to be as well as how it was taught (National Council of Teachers of Mathematics, 1989; National Research Council, 1989). Mike Smith and Jennifer O’Day (1990) wrote an iconic piece describing their vision of “systemic school reform,” which in contrast to local school restructuring reforms would build out new, content-driven state structures. Clear and challenging standards for student learning would provide an organizing framework toward which other policy tools could be focused. Lauren Resnick sat on a dozen policy commissions and taught politicians and business leaders about the cognitive science behind the “thinking curriculum.” She also helped explain why the decomposability and decontextualization assumptions implicit in the machinery of standardized tests were inimical to teaching for deep understanding (Resnick & Resnick, 1992).

Among many policy documents, a report of the NAEd (McLaughlin & Shepard, 1995) focused in particular on the learning theory, assessment, and equity arguments underlying standards-based reforms. An immense body of research from the 1980s and 1990s, later codified in *How People Learn* (National Research Council, 1999a), included new insights

about the nature of expertise, knowledge structures and connections to prior knowledge, the importance of metacognitive strategies and self-regulation, links between motivation and sense of self to what is learned, and even the emerging idea (then) that participation in social practices is a fundamental aspect of learning. Examples from this research base are myriad. Studies of learning in out-of-school settings, such as Collins, Brown, and Newman's (1989) cognitive apprenticeship, demonstrated the importance of situating abstract tasks in authentic contexts, very different from the inert and decontextualized forms of knowing required in schools. Lampert (1990) sought to shift classroom participation structures to more closely resemble standards of logical argument in the mathematical community. Cobb, Wood, and Yackel (1993) argued that new norms would need to be negotiated to overcome previously constructed norms about trying to guess the teacher's solution and avoiding evaluation. Luis Moll's "funds of knowledge" for teaching offered a practice whereby students' prior knowledge about agriculture, mining, economics, household management, medicine, and religion could be used as cognitive resources to engage students in more challenging and meaningful tasks (Moll, Amanti, Neff, & Gonzales, 1992).

Although the above ideas were widely shared among researchers, the NAEd report tried to explain to a popular audience what findings from cognitive and constructivist psychology meant for changes in classroom practice. For example, it was a mistake, left over from behaviorism, to postpone thinking and reasoning until after basic skills were learned by rote. The NAEd report also discredited widely held nativist beliefs about inherited abilities that lurked behind contemporary instructional practices, making it acceptable to reserve rich and engaging curricula for only an elite subgroup of students. Thus, the NAEd report endorsed, in principle, the idea of "high standards for all students" but noted that the needed changes to the system were unprecedented and monumental. The report cautioned further that despite hopes for greater opportunity and equity, reforms could actually exacerbate inequities if standards were accompanied by high-stakes assessments, if teachers in urban and poor school systems had the least access to professional development, and if students were punished for failures that were the system's fault. The authors tried to explain the apparent contradiction of knowing a great deal about learning and teaching but not having sufficient knowledge about how larger systems and social contexts could be sufficiently transformed to make the envisioned changes possible. The report emphasized the importance of capacity building, especially teacher professional development, and the need for continuous research and evaluation of the reforms' effects.

How naïve it was to imagine that policymakers' past practices would not trump ephemeral visions of reform. In 1994, two different versions of

standards-based reforms were installed in federal policy, Clinton's Goals 2000: Educate America Act and the reauthorization of the Elementary and Secondary Education Act (ESEA) called the Improving America's Schools Act (IASA). Both alluded to the systemic changes that would be needed, but it would be up to states to figure out how to make and fund those changes. IASA included principal and teacher professional development as part of Title II; but it was IASA's accountability mandates that determined its character and impacts. Subject-matter experts kept talking about research on learning, but when policymakers adopted rewards and sanctions as the drivers of change, standards-based reforms became an incentives theory of change. Mathematics education reformers developed beautiful examples of curricular resources that would help teachers help students develop deeper conceptual understandings; and learning-focused projects such as TERC and LRDC's Institute for Learning developed resources that could support transformative change. But none of these could have the reach of accountability mandates, which by definition affected every classroom and school.

By 1999 a National Research Council (NRC) report titled *Testing, Teaching, and Learning* (National Research Council, 1999b) concluded that the theory of action guiding standards-based reform might be overly optimistic because it assumed that teachers know how to educate all children to high levels of performance or know how to seek the relevant new knowledge. Accountability structures were thought to be sufficient to bring these extant resources to bear. Studies of what was happening on the ground, however, found that external accountability mandates landed in very different ways in rich and poor schools. Better-situated schools, as termed by Carnoy, Elmore, and Siskin (2003), that is, those serving more advantaged communities, were more able to respond coherently to accountability pressures. Better-positioned schools with relatively high "internal accountability" recognize that increased coherence around instructional practice required new curriculum content and new knowledge and skills for teachers and administrators—and found ways to change the structure of the work and gain those skills. Without this wherewithal, the reforms were a hallow shell.

The assessment strand of standards-based reforms has a similar history of grand hopes followed by erosion and misdirection. Evidence gathered in the late 1980s showed the negative effects of teaching to low-level tests. In particular, an important synthesis project led by Michael Feuer for the Congressional Office of Technology Assessment (U.S. Congress, 1992) brought together studies documenting the curriculum narrowing effects of high-stakes testing and resulting test score inflation, that is, test scores could go up without there being a corresponding increase in learning. In the United States, performance assessments were offered by

standards-based reformers as the remedy to these distorting effects. New assessments that better represented nobler learning targets would be “tests worth teaching to.” In England and other countries in the United Kingdom, an Assessment Reform Group focused instead on formative assessments in classrooms as a potential counterforce to the damaging effects of school league tables. Building specifically on the important contributions of *How People Learn*, an NRC committee was formed to bring together new knowledge from research on both learning and measurement. The resulting NRC report, *Knowing What Students Know* (National Research Council, 2001), explained the different purposes of large-scale versus classroom-level assessments and how the two could be coherently linked by a shared model of learning.

From this foundational knowledge, the 1990s saw a brief flourishing of more inventive forms of assessment. These included portfolio assessments in Kentucky and Vermont and performance assessments in California and Maryland. But this heyday was short lived. Perhaps the most visible example was the California Learning Assessment System (CLAS) that lasted only 3 years. Religious groups objected to the content of reading passages and to the idea of the Sacramento bureaucrats assessing “critical thinking;” and newly elected policymakers resented the tradeoff that had been made, sacrificing individual student scores to make performance assessments possible (Kirst & Mazzeo, 1996). The real death knell to performance assessment reforms, however, came with the passage of the No Child Left Behind Act (NCLB) in 2001. The sheer volume of tests required and the mandate for individual student scores closed down any state testing program that had relied on matrix sampling to obtain school scores and made scoring of open-ended assessments cost prohibitive. NCLB required that every child be tested every year in reading and math from grades 3 to 8. Moreover schools would essentially be placed in receivership if they failed to demonstrate adequate yearly progress defined by increasingly out-of-reach targets. The idea that 100 percent of students would be expected to meet ambitious learning targets by 2014 was absurd on the face of it. By 2011, states began receiving waivers from some of the more draconian requirements, but this did not prevent a frantic, decade-long focus on raising test scores. NCLB also had an explicit focus on closing gaps between majority and minority groups, but its provisions failed to attend to the kinds of genuine learning opportunities that would make these leaps possible.

NCLB’s relentless accountability pressures had further pernicious effects because of what the focus on test scores did to undermine the fledgling efforts being made to introduce formative assessment practices in the United States. In my address to the American Educational Research Association in 2000 titled “The Role of Assessment in a Learning Culture,”

I took up the formative assessment arguments of colleagues in the United Kingdom, Australia, and New Zealand and tried to draw connections between learning processes described from a Vygotskian perspective and what subject-matter experts were saying about ambitious learning goals. Numerous learning principles rendered from a cognitive perspective in *How People Learn*—attending to prior knowledge, substantive feedback, internalizing criteria, metacognitive benefits of self-assessments, teaching, and assessing for transfer—can also be taken up in socially supporting learning environments in ways that enable collaborative relationships between student and teacher (Gipps, 1999). In such a culture, developing an identity of mastery occurs as learners participate in a community of practice (Lave & Wenger, 1991). But these ideas cannot flourish in a test-driven environment.

In the wake of NCLB, entrepreneurs and test publishers co-opted the term formative assessment and used it to sell products to school districts with item formats that were cheap imitations of state tests. Another reform was undone by superficial understandings. Dylan Wiliam (personal communication, 2005) called these products “early warning summative tests.” In essence, districts were paying good money for instruments that looked for all the world like teaching-the-test training materials. Patricia Burch (2006), who studies various forms of educational privatization, found that top testing vendors doubled their annual sales between 2000 and 2006. A group of scholars brought together under the auspices of the Council of Chief State School Officers issued a formal decree explaining why more frequent testing with mostly multiple choice items bore no resemblance to the learning research supporting formative assessment. The very tiniest victory was won when the term interim (Perie, Marion, & Gong, 2009) or benchmark assessments was adopted instead of formative assessments to describe formal tests given every 4 to 6 weeks. However, the use of interim assessments themselves in no way abated.

Simply examining the most popular of these commercial test products should have made it clear why they are so unlikely to produce deep and meaningful changes in learning opportunities. For the most part, though computer delivered, they look just like the narrow basic skills tests from the 1980s. They were not developed to provide diagnostic insights about students’ understandings. Empirical studies examining the use of such measures show that instructional responses are largely procedural or at best they only prompt teachers to try something different (Nabors Olah, Lawrence, & Riggan, 2010). The few positive examples of benchmark assessment results being more deeply linked to instruction improvements appear to be led by highly committed principals or teacher leaders, but the more prevalent practice is item-by-item reteaching with little attention to student thinking (Blanc, Christman, Liu, Mitchell, Travers, & Bulkley,

2010; Shepard, 2010; Shepard, Davidson, & Bowman, 2011). In our study of interim assessments in seven districts in two states, we found that teachers described not a learning culture but a benchmark assessment or accountability culture exemplified by posting students' scores in the hallway and giving feedback to students in terms of how many more items they needed to score correctly to reach proficiency (Shepard et al., 2011).

The NAEd white paper on *Standards, Assessments, and Accountability* (Shepard, Hannaway, & Baker, 2009) included some of this same history on standards-based reforms as well as a summary of policy research documenting limited investments in capacity building. Effective examples of teacher professional development were cited where they did occur. The NAEd working group authors made several recommendations about needed changes within the existing federal accountability framework, calling for well-articulated learning progressions, ambitious but realistic learning targets, ongoing evaluation of accountability systems to ensure that they contribute to school improvement, and closer investigation of school performance before imposing remedies or sanctions. In my view, however, the most important of our recommendations was the following: "The federal government should support an intensive program of research and development to create the next generation of performance assessments explicitly linked to well-designed content standards and curricula."

In this one recommendation are two critically important ideas: first, that an intensive program of research is needed to develop next-generation assessments and second, that performance assessments and curricula should be developed together.

Given the theme of this essay—that reforms are undone by superficial understandings or by hollow enactments of idealized schemes—it should not be surprising that the idea of an intensive assessment research and development (R&D) effort was undermined, essentially by the decision to deliver new operational tests on a broad scale in too short a time. The Department of Education heard the argument that state consortia would be needed to build and try out the kind of curriculum-linked, learning-progression-linked assessments outlined in the NAEd white paper and in *Knowing What Students Know* a decade before. They understood that individual states would not be able to undertake such challenging development on their own. However, the distinction we had drawn between "the political process needed to achieve consensus and guide policy decisions versus the scientific expertise needed to develop and rigorously evaluate curricular materials, instructional strategies, and assessments" (Shepard, Hannaway, & Baker, 2009, p. 8) was lost.

Following the Great Recession, the American Recovery and Reinvestment Act monies made it possible for the federal government to invest in

developing next-generation assessments. Expert testimony was sought, which forewarned of all the past problems, but sometimes promised grandly how these problems would be overcome by the affordances of technology. The resulting Race to the Top call to develop Comprehensive Assessment Systems (U.S. Department of Education, 2010) was layered with enormous demands requiring that consortia comprising at least 15 states develop measures of college- and career-ready, cross-grade achievement trajectories in partnership with higher education institutions. Proposers were expected to correct all of the shortcomings of past assessments: measure the full range of performance implied by the standards, including aspects of achievement that had heretofore been difficult to measure; elicit complex applications of knowledge and skills; measure accurately for high- and low-achieving students; and so forth. The successful consortia were also required to ensure that their assessments were “valid, reliable, and fair.” Professionals involved in the consortia performed herculean tasks, but there was no way to live up to all of the rhetorical claims, and some shortcuts were necessary. For example, validity analyses had to rely more on content reviews by experts and internal statistical properties of the assessments during field trials rather than empirical studies of assessment results linked to either school improvement or student success in college and career.

The two consortium tests, Partnership for Assessment of Readiness for College and Careers (PARCC) and Smarter Balanced, were rolled out in spring of 2015. It is fair to say that these new assessments are generally of higher quality than past state assessments because they are targeted at higher levels of thinking and use more open-ended formats. They are not so good, however, as what might have been possible if investments had been made in a genuine R&D effort. It is also fair to say that both consortia have made some significant missteps, assuming too much about computer availability in all schools, requiring too much testing time, and sometimes using drag and drop and other technological interfaces in ways that hurt rather than enhance representations of important content. Consortium leaders are still trying to solve too many irreconcilable problems. Broad coverage with open-ended performance assessments would be possible with reasonable amounts of testing time if policymakers would reconsider the possibility of matrix sampling and roll back the demand for test-based teacher accountability. More importantly, policymakers might also recall, from the long history of standards-based reforms, that top-down mandates do not help poor schools get better if educators in these schools do not have access to resources to teach in fundamentally different ways.

In the fall of 2015, just before this essay is to be published, results will be released showing miserable student performance on PARCC and Smarter Balanced. No amount of explaining will help the public under-

stand the ambitions of the new content and practices frameworks, the stringency of proficiency cutoffs, nor the lack of resources or time to turn an entire system toward these new ends. Public schools will be bashed again and dedicated, caring teachers will continue to leave the profession in droves.

I began this account by asking why educational reforms are not successful, given that we know so much from research on learning. The answer and lesson to be learned by researchers as well as policymakers is that *cheap, superficial, and coercive versions of reform ideals will inevitably prevent deeply substantive, hoped-for changes*. The kinds of transformative changes that are needed—to make real differences in learning opportunities—are difficult and cannot be made on command. No amount of talk about “capacity building” can substitute for the supports that are needed. As predicted, inequities are increased when short-cut strategies are the best that can be done in response to accountability pressures. Researchers who helped conceptualize the beginning of the standards movement in the 1990s wanted to create policy coherence at the top that would support meaningful changes at the bottom of the system. But oppressive regimes at the top only create scurrying at the ground level. That is why drilling—on worksheets or interim measures that imitate accountability tests—has been so much more pervasive than deeper changes in curriculum or instructional practices.

We are now in a world of next-generation, Common Core State Standards (or new state standards that closely resemble CCSS), and goals such as critical thinking and problem solving now have wider appeal. If policy leaders want to support more profound changes in teaching and learning opportunities—in poor as well as rich schools—then they will need to reconsider the juggernaut of accountability testing. To do this it might be helpful to return to the recommendations from *Knowing What Students Know* and recall the very different purposes of large-scale assessments for monitoring and accountability versus classroom-level tests to inform teaching and learning. There will surely need to be refinements of PARCC, Smarter Balanced, and various other state tests. Ideally they would be used to collect data and track progress, not to create incentives by imposing unreasonable targets. If leaders insist on targets, then they should be informed by what Bob Linn (2003) called an “existence proof,” that is, high standards that at least someone has been able to reach; for example, schools might be asked to raise achievement to the levels currently attained by the top 25 or 10 percent of similarly situated schools.

Not to be forgotten, an important and distinctly different need is for the development of curriculum materials to support teachers in learning to teach to much more ambitious standards. The design of assessment tasks, both formative and summative, should be an integral part of

such curriculum development. The National Science Foundation's current funding of learning progressions in science is one way to study jointly the furthering of learning at the same time that we get better at assessing and interpreting student thinking. If we have learned only one thing from the disappointments of standards based reforms, then it should be that trying to install giant systems is a mistake. It would be much better to take a step back from the most aversive aspects of current accountability systems and focus instead on smaller scale projects with adequate time to learn from mistakes and improve. Then we could imagine implementing such curricular materials and next-generation assessments on a larger scale, but only if at each step we have evidence that systems are becoming more equitable, not less.

References

- Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: Benchmarks and instructional communities. *Peabody Journal of Education*, 85(2), 204–225.
- Burch, P. (2006). The new educational privatization: Educational contracting in the era of high stakes accountability. *Teachers College Record*, 88(2), 129–135.
- Carnoy, M., Elmore, R., & Siskin, L. S. (2003). *The new accountability: High schools and high-stakes testing*. New York: Routledge Falmer.
- Cobb, P., Wood, T., & Yackel, E. (1993). Discourse, mathematical thinking, and classroom practice. In E. A. Forman, N. Minick, & C. A. Stone (Eds.), *Contexts for learning: Sociocultural dynamics in children's development* (pp. 91–119). New York: Oxford University Press.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Erlbaum.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. In P. D. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education* (Vol. 24, pp. 355–392). Washington, DC: American Educational Research Association.
- Goals 2000: Educate America Act. 20 U.S.C. § 5801, et seq. (1994).
- Improving America's Schools Act, 20 U.S.C. § 6301, et seq. (1994).
- Kirst, M. W., & Mazzeo, C. (1996, April). *The rise, fall and rise of state assessment in California, 1993–1996*, paper presented at the annual meeting of the American Educational Research Association, New York.
- Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, 27(1), 29–63.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.
- Linn, R. L. (2003). Performance standards: Utility for different uses of assessments. *Education Policy Analysis Archives*, 11(31). Retrieved from <http://epaa.asu.edu/epaa/v11n31>.
- McLaughlin, M. W., & Shepard, L. A. (1995). *Improving education through standards-based reform. A report of the National Academy of Education Panel on Standards-Based Reform*. Stanford, CA: National Academy of Education.

- Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms, *Theory Into Practice*, 31(2), 132–141.
- Nabors Olah, L., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results, *Peabody Journal of Education*, 85(2), 226–245.
- National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.
- National Research Council. (1989). *Everybody counts: A report to the nation on the future of mathematics education*. Washington, DC: National Academy Press.
- National Research Council. (1999a). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academy Press.
- National Research Council. (1999b). *Testing, teaching, and learning: A guide for states and school districts*. Washington, DC: National Academy Press.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Perie, M., Marion, S., & Gong, B. (2009). Moving toward a comprehensive assessment system: A framework for considering interim assessments. *Educational Measurement: Issues and Practice*, 28(3), 5–13.
- Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for education reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (pp. 37–75). Boston, MA: Kluwer Academic.
- Shepard, L. A. (2010). What the marketplace has brought us: Item-by-item teaching with little instructional insight. *Peabody Journal of Education*, 85(2), 246–257.
- Shepard, L., Hannaway, J., & Baker, E. (Eds.). (2009). *Standards, assessments, and accountability*. Washington, DC: National Academy of Education.
- Shepard, L. A., Davidson, K. L., & Bowman, R. (2011). *How middle-school mathematics teachers use interim and benchmark assessment data*. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Smith, M. S., & O'Day, J. (1990). Systemic school reform. In S. H. Fuhrman & B. Mahen (Eds.), *The politics of curriculum and testing: The 1990 yearbook of the Politics of Education Association* (pp. 233–267). London, UK: Taylor & Francis.
- U.S. Congress, Office of Technology Assessment. (1992). *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Government Printing Office.
- U.S. Department of Education. (2010). Overview information: Race to the Top Fund Assessment Program; Notice inviting applications for new awards for fiscal year (FY) 2010. *Federal Register*, 75(68), 18171–18185.

The National Academy of Education
and the Professional Development
of Education Scholars

Celebrating Our Successes and Offering a Cautionary Note: Reflections on the National Academy of Education/ Spencer Fellowship Programs

Hilda Borko and Lois Weis¹

The National Academy of Education (NAEd), together with the Spencer Foundation, supports two very prestigious and respected fellowship programs for early career scholars conducting research in the field of education. As Michael McPherson, president of the Spencer Foundation, remarked, the NAEd/Spencer Foundation Postdoctoral and Dissertation Fellowship Programs “have been the Foundation’s signature programs for most of its nearly 40-year lifetime.... Many of the leading lights in the study of education are themselves past recipients of either or both of these fellowship awards” (The Spencer Foundation, 2011, p. 3). In this essay we reflect on the value of these two flagship programs and raise some cautionary issues, drawing from evaluations sponsored by the Spencer Foundation, as well as our experiences on the selection committees for the two programs and the fellowship Retreat Planning Committee.

Hedges and Asch (2010) note

The Spencer Foundation Fellowship Program² was established to enhance the quality of educational research by improving the human infra-

¹ Hilda Borko is Professor of Education at Stanford University. She was elected to the National Academy of Education in 2005. Lois Weis is State University of New York Distinguished Professor at the University at Buffalo, State University of New York. She was elected to the National Academy of Education in 2013. The two authors contributed equally to this article and listed their names alphabetically.

² Like the postdoctoral program, the dissertation program is now administered by the NAEd with support from Spencer. It is now formally known as the National Academy of Education/Spencer Foundation Dissertation Fellowship Program.

structure for educational research. In order to accomplish this goal, the fellowship program provides dissertation fellowships annually to 25–30 young scholars who are writing doctoral dissertations on education-related topics in various fields. The fellowships provide support for doctoral students to carry out their dissertation research, thus advancing their research careers. (p. 1)

The NAEd/Spencer Postdoctoral Fellowship Program has the same central goal. The fellowships, awarded annually to 25–30 scholars who are within 5 years of receiving their doctorates, afford recipients the opportunity to carry out substantial education research projects, “thus advancing their research agenda and their careers” (Hedges & Hanis, 2005, p. 2).

The two programs support the fellows through a combination of significant financial support and carefully planned retreats. The retreats provide an opportunity for fellows in the two programs to interact and learn from one another, meet with leading education scholars for focused interactions around their work, and participate in discussions led by senior scholars that address scholarly and professional issues relevant to early career scholars.

Embarking upon evaluation of the impact of these Programs, Hedges and colleagues set out to determine the extent to which fellows had more successful careers as a result of the fellowship. As they explain,

The principal analytic problem is to first evaluate how much better we should expect the treatment group to be if there were no treatment effects, and then determine whether the treatment group is better than would be expected on this basis. The treatment effect is the difference between how much better we would expect the treatment group to be (if there were no treatment effect) and how much better it actually *is*. (Hedges and Asch, 2010, p. 2)

Both evaluations employed a regression discontinuity design to estimate the effects of fellowship receipt on outcomes of interest (comparing outcomes for fellowship recipients with those of finalists who did not receive the fellowship, adjusted for the promise exhibited prior to the fellowship). Outcome measures were also similar. For both evaluations, they included research productivity, individual’s influence on others’ research, and measured ability to garner grants/fellowships to support their research. In addition, the evaluations of both fellowship programs included the individual’s academic rank. The evaluation of the dissertation fellowship program included completion of the doctorate.

Hedges and Asch conclude that receipt of a dissertation fellowship enhanced the probability of the following: completion of the PhD, key measures of research productivity, measured influence on the work of others, and number of federal and non-federal grants received. In addition,

receipt of a dissertation fellowship had a small “but statistically insignificant effect on the likelihood of holding an academic ladder rank position” (p. 14). Results of the postdoctoral program evaluation were similar. Participation in the program had positive effects on all four measured dimensions of academic career success. Results were statistically significant for research productivity (as measured by total number of publications) and ability to garner resources to support research (as measured by number of research grants obtained) (Hedges & Hanis, 2005).

Receipt of a Spencer Fellowship seemingly “turbo charges” individuals with regard to particular kinds of career criteria, the very career criteria that are arguably becoming more and more important at the point of hiring, promotion, and tenure. As Hedges, Hanis, and Asch (2011) observe, however, although the program evaluations “help establish that the fellowship programs have positive effects, they shed little light on the specific mechanism(s) that produce(s) those effects” (p. 25). These authors suggest that outcomes are likely affected by a combination of financial, programmatic, and mentorship aspects of the two programs, and that the magnitude of the effects may differ across the two programs and the various outcomes. Our extensive combined experiences on the two Fellowship Selection Committees (7 years), Professional Development Committee (4 years), and Retreat Planning Committee (3 years) suggest that this may indeed be the case, and that further study is warranted to better understand the impact of the two programs on fellows’ careers and the mechanisms through which such impact is exerted.

We also anticipate that the programs’ short- and long-term impact on recipients will only become stronger as a result of the NAEd’s increased attention to and support of their programmatic and mentorship components. For example, in 2011 a committee was appointed to plan the retreats for the postdoctoral and dissertation fellows, distinct from the selection committee (rather than having the selection committee also be responsible for planning and leading the retreats), thereby acknowledging the labor intensity of the work and enabling each committee to devote its time and energy to a single program component.

Building on and broadening this recommendation, in 2011 the Board of Directors formed the Professional Development Committee (PDC) in order to have one overarching governance structure that would provide conceptual leadership for all professional development activities, and would oversee the strategic direction of the NAEd/Spencer Postdoctoral Fellowship Program, the NAEd/Spencer Dissertation Fellowship Program, and other constituent groups such as those comprised of former fellows. The PDC also monitors and evaluates each program’s progress, and makes recommendations for the appointment of individuals to the two fellowship selection committees and the retreat planning committee.

In an effort to further strengthen programmatic and mentoring opportunities for the fellows, the PDC created a single Retreat Planning Committee charged with planning and orchestrating semiannual retreats for the combined group of dissertation and postdoctoral fellows (rather than separate committees and separate retreats). That committee is chaired by the PDC chair and includes two additional PDC members and one former recipient of each of the fellowships. This consolidated committee has achieved a truly integrated and intergenerational retreat, with increased participation of former fellows in retreat activities. Although no data are available, it is reasonable to assume that these changes have increased, and will continue to increase, the networking opportunities embedded in the fellowship programs and that they are additionally helping to build and strengthen a scholarly community of current and former fellows.

The above noted positive impact of the fellowship programs also leads to some concerns about potential bias in the selection process; particularly, in this case, the extent to which candidates from *particular universities* are privileged with regard to fellowship receipt.³ We express this concern because it is arguably the case that findings related to the “effects” of the two programs are intellectually parallel, or perhaps metaphorically parallel, to the growing body of work by scholars such as Bowen and Bok (1998); Bowen, Chingos, and McPherson (2009); and Stephan, Rosenbaum, and Person (2009), who suggest that *where* one goes to college predicts persistence and graduation above and beyond entering characteristics of admitted students. Evidence also suggests that selective institutions are better resourced than less selective institutions and confer on their graduates both special entrée to the best graduate and professional programs in the country (Eide, Brewer, & Ehrenberg, 1998) and well-documented labor market advantages (Bowen & Bok, 1998; Rumberger & Thomas, 1993; Thomas, 2000; Thomas & Zhang, 2005). As with persistence and graduation rates, these relationships hold even when characteristics of entering students are held constant in the analysis.

To be clear, highly selective institutions are extraordinary places, and admitted students work very hard to gain access to these schools, persist, and graduate. Having said this, anyone who has anything to do with this group of institutions knows that opportunities to create future success are deeply embedded within the institutions themselves, and that measured outcomes of these institutions are not simply a matter of “signaling.” Yale

³ We express this concern for both the doctoral and postdoctoral programs, although the issue is not exactly the same for the two groups of fellows. With respect to the doctoral program, we are concerned about advantages accrued by virtue of current graduate work in particular institutions. For postdoctoral fellows (who have taken up academic positions in a wider range of institutions than where they did their degrees), we are concerned about overrepresentation of institutions from which they received their doctoral degree.

is Yale to be sure, but the mechanisms embedded within schools such as Yale to build a dossier that lays the foundation for one's future graduate and professional studies and ultimately one's career, are notable. For example, well-funded not-for-profit programs affiliated with particular campuses, such as Princeton in Africa, enable recent graduates to work in South Africa or other nations for a year. In so doing, postgraduates gain important skills and cultural and social capital linked to being a world citizen, and they additionally have something to write about in subsequent law, medical, graduate school or other admissions essays. Such programs offer a pathway to a level of future professional engagement and position that goes well beyond institutional "signaling" of worth.⁴ There are actual mechanisms in place wherein students who attend these institutions are able to persist, graduate, and gain opportunities for notable dossier building experiences to a greater extent than students who attend less privileged institutions. These mechanisms enable graduates to *distinguish* themselves as they position for access at the next level, thereby systematically scaffolding future possibilities.

Our point here is that receiving an NAEd/Spencer dissertation or postdoctoral grant is comparable, and while all those who receive prestigious fellowships definitely work hard and certainly deserve every bit of social and economic capital that accrues as a result of these prestigious programs, those of us who sit on the NAEd/Spencer Dissertation and Postdoctoral Fellowship selection committees must continually ask ourselves "how do we 'create a class'?"⁵ What are the criteria through which a "class" in any given year is created—a "class" that we know, based on the data presented earlier, turbocharges a group of people while simultaneously accomplishing exactly what the Spencer Foundation or any foundation would want out of its funding. It is a very good investment of available funds.

In our collective years on the dissertation and postdoctoral selection committees, we know firsthand that the selection committees are very thoughtful about this issue, and do not simply revert to institutional sig-

⁴ Although Princeton in Africa has recently been opened to graduates of other institutions, we can assume that a significant portion of the fellowships will continue to go to graduates from Princeton. Although we highlight Princeton in Africa, programs aimed at offering a myriad of postgraduate opportunities are routinely available at particularly located universities. For example, the Project 55 Fellowship program "connects talented Princeton graduates with some of the most effective and innovative public interest organizations around the United States" (https://home.alumni-corp.org/?page_id=127). All such fellowships scaffold future educational and job market possibilities.

⁵ We use the term "class" not to designate social class, but rather, in the same way institutions use the term when they put together a law school class or class for purposes of freshman admission.

naling when making awards. The problem is that some institutions are far “thicker” than others with respect to the human and material capital that can be collectively deployed toward mentoring and supporting candidates’ applications. While outstanding researchers take up positions at a range of institutions, the best endowed of such institutions have well-formed collectivities in one or more identifiable scholarly areas. This “richness” enables applicants from particularly located institutions to be far more competitive than those who do doctoral work at less prestigious universities.⁶

By way of example, knowledge is not equally distributed among faculty (in this case, between varying institutional locations) with regard to what constitutes a “winning” letter of recommendation, and applicants who study with faculty who possess such knowledge are at a clear advantage in the selection process. Letters count—and they count a lot—and certain faculty know, by virtue of their own located position in the structure of opportunities, how to write them.

Similarly, individually constructed dossiers count in the selection process—and they count a lot—and applicants who have attended a relatively small stratum of institutions (as doctoral students) are provided the kind of advice on a routine basis that they must have in order to put forth a dossier that is at least competitive, if not winning. Others cannot and do not, although they may be equally as smart, and, at some level, equally or even more accomplished given possibilities available in their graduate programs relative to those in more well-endowed programs. In point of fact, the competition for these fellowships is fierce, and a high proportion of award winners enter the competition with a corpus of publications (at times in tier one journals) and an already established record of grant work by virtue of their current or former association with grant-worthy faculty at a relatively small number of top-tier schools.

The NAEd and Spencer Foundation are very concerned about these issues of access and unintentional privileges and have taken a number of important steps to address them. Both organizations advertise the fellowship programs broadly, and both websites provide links to extremely informative webinars for all interested individuals. Individuals at any degree-granting institution—faculty or student—can theoretically access

⁶ Looking at dissertation fellowship recipients across the years 2010, 2011, 2012, and 2013, for example, we see that if we use three or more grantees as our criterion over a 4-year period, dissertation fellows come from a very narrow band of schools—Harvard (10); Stanford (10); Berkeley (11); Northwestern (3); University of Chicago (3); TC Columbia (3); U. Michigan (3); and Penn (3). If we up the criterion to more than 10 awardees over this same time period, fellows come from three institutions—Harvard, Stanford, and Berkeley. This means that of the 100 total awards, almost one-half are awarded to individuals who come from eight schools, and close to one-third come from three schools.

these excellent webinars and position themselves for fellowships. In addition, the PDC recently offered a well-attended session at the American Educational Research Association whose goal was to provide increased access to information about the programs to a broadened range of potential applicants.⁷

Members of the selection committees are cognizant of these issues and are well aware that access, in and of itself, does not translate into equal opportunities. We continue to think carefully about ways to democratize access for fellowships without compromising quality. Given the commitment of both the NAEd and the Spencer Foundation to broadening access for historically marginalized populations, we attempt to take seriously the point that facially neutral policies and practices that do not take into account differential access to particular kinds of resources among differentially located classes of people will not make a dent in the power of institutional inequalities with regard to outcomes noted here.

We conclude this essay with both kudos and a challenge. The NAEd/Spencer Dissertation and Postdoctoral Fellowships are one of the NAEd's most important and lasting contributions to educational scholarship. We applaud recent efforts to increase the number of fellowships awarded and strengthen key mentorship and community-building components. At the same time, the NAEd members must continue to ask ourselves: do we perhaps unintentionally perpetuate and increasingly exacerbate institutionally linked inequalities with policies and procedures that do not adequately take into account that a particular "class" of students is negatively impacted by virtue of prior background and position within the structure of opportunities that is itself now intensifying in higher education?⁸ In addition we must continue to identify ways of making it easier for at least some young scholars who are in less privileged institutions to become competitive for these highly impactful programs.

References

- Bowen, W. G., & Bok, D. (1998). *The shape of the river: Long-term consequences of considering race in college and university admissions*. Princeton, NJ: Princeton University Press.
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at America's public universities*. Princeton, NJ: Princeton University Press.

⁷ This session drew a very large group of faculty and students from a range of postsecondary institutions. These kinds of efforts signify the NAEd and Spencer Foundation's concern about issues expressed in this chapter.

⁸ Here we use class in the sense of background characteristics, including in this case experiences in particularly located postsecondary institutions at the undergraduate and graduate level.

- Eide, E., Brewer, D., & Ehrenberg, R. (1998). Does it pay to attend an elite private college? Evidence of the effect of undergraduate college quality on graduate school attendance. *Economics of Education Review*, 17, 71–376.
- Hedges, L., & Asch, E. (2010). *Final report on the evaluation of the Spencer Dissertation Fellowship Program*. Chicago, IL: The Spencer Foundation.
- Hedges, L., & Hanis, J. (2005). *Report on the evaluation of the NAE/Spencer Postdoctoral Fellowship Program*. Chicago, IL: The Spencer Foundation.
- Hedges, L., Hanis, J., & Asch, E. (2011). Statistical evaluations of the Spencer fellowship programs. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part II, pp. 9–26). Chicago, IL: The Spencer Foundation.
- McPherson, M. (2011). Introduction. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part 1, pp. 2–8). Chicago, IL: The Spencer Foundation.
- The Spencer Foundation (2011). *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs*. Chicago, IL: The Spencer Foundation.
- Stephan, J., Rosenbaum, J., & Person, A. (2009). Stratification in college entry and completion. *Social Science Research*, 38(3), 572–593.
- Rumberger, R.W., & S. L. Thomas. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review*, 12, 1–19.
- Thomas, S. (2000). Deferred costs and economic returns to college quality, major and academic performance: An analysis of recent graduates in baccalaureate and beyond. *Research in Higher Education*, 44(3), 263–299.
- Thomas, S. L., & Zhang, L. (2005). Changing rates of return to college quality and academic rigor in the United States. Who gets good jobs in America? *Research in Higher Education*, 46, 437–459.

Professional Development for Early-Career Scholars

*Marilyn Cochran-Smith*¹

In the spring of 2011 under the presidency of Susan Fuhrman and then with strong support under the current presidency of Michael Feuer, the National Academy of Education (NAEd) instituted a Professional Development Committee (PDC), which I had the privilege of chairing for 4 years. Identifying and supporting the development of the next generation of education scholars was not a new mission for the NAEd, which had sponsored a postdoctoral fellowship program for years before that. What prompted the formation of a committee specifically focused on professional development at this moment in time was the NAEd's new responsibility for administering the Spencer Foundation's dissertation fellowship program in addition to the ongoing postdoctoral fellowship program. This addition to the NAEd's portfolio of projects was regarded as an unprecedented opportunity to develop a unified professional development framework for identifying and nurturing the scholarship and career development of a new generation of education researchers at a critical stage of their training.

Twice-yearly the NAEd retreats for dissertation and postdoctoral fellows became a kind of crucible for the Academy's emerging professional development agenda. It was in this context that many emerging ideas and approaches for nurturing the work of young education scholars were tried out, adapted, accepted, resisted, or ignored. The Academy's cur-

¹ Marilyn Cochran-Smith is the Cawthorne Professor of Teacher Education for Urban Schools at Boston College. She was elected to the National Academy of Education in 2009.

rent approach to professional development emerged over time from the iterative retreat planning and evaluation process coupled with thoughtful ongoing discussions of the NAEd Board, the PDC, and the two NAEd fellowship program selection committees. The members of these selection committees take their responsibilities very seriously and devote significant time and energy to this worthy process.

The most prominent feature of the Academy's unified professional development approach is that it is both generational and intergenerational. That is, we recognize that there are tasks, issues, dilemmas, questions, and needs shared by education scholars who are at particular career phases (e.g., doctoral students working out dissertation analyses, assistant professors in the first few years on the job, just-promoted associate professors who have new responsibilities) or at different stages of scholarly development (e.g., envisioning a research agenda that goes beyond dissertation work and/or collaboration with a mentor, branching out to include a new methodology that strengthens an already well-developed research topic) and thus are part of particular "generational" groups. The unique tasks and demands each group faces unite the generations and also set them apart from other generations who are dealing with different demands. This means that each generation of scholars needs professional development that addresses its own needs.

On the other hand, we also recognize that there is enormous benefit to professional development wherein scholars from different generations learn from and with those who are behind or beyond them in terms of career phase or scholarly development. The Academy's unified approach to professional development accommodates the needs of each generation of scholars, but it concentrates on intergenerational learning opportunities. These are at the heart of twice-yearly fellows retreats where dissertation fellows and postdoctoral fellows come together with former fellows and senior scholars to teach and learn from each other. Accordingly, the focus of sessions during the retreats is both professional and scholarly. For example, at each retreat, fellows select sessions related to pressing professional issues, such as job searches/job talks, the tenure process, the transition from assistant to associate professor, collaborating across disciplines, competing for grants, and getting one's work published in top journals or by prestigious university presses. The retreats also feature many opportunities to concentrate on the substance of each individual's scholarship through mentoring sessions that feature a senior scholar, often specifically requested by fellows, and a small group of both dissertation and postdoctoral fellows who are linked by disciplinary field, methodology, or topic. In addition the retreats always have invited speakers who speak to current educational issues and policies, often presenting cutting-edge work that is thought-provoking for everyone, regardless of generation.

Two recent retreat activities illustrate the intergenerational approach. During the fall retreat, which overlaps with the annual meeting of the NAE, second-year postdoctoral fellows present their research during a series of concurrent “fellows forums,” each of which features 3–4 presentations with a senior NAE discussant who provides integrative comments. Dissertation fellows, first-year postdoctoral fellows, and the NAE members who are present for the annual meeting are all part of the audience. Scholars from each of these generational groups actively participate in the question-answer sessions, and all of them have opportunities to interact informally throughout the meeting. These formal and informal settings provide young scholars with many opportunities to have their work validated through the interest and questions of more senior colleagues, and many lasting collegial relationships are formed during these events.

A second example of the intergenerational professional development approach comes from a recent spring retreat, which is exclusively for fellows and invited senior scholars. During this retreat, the main event is the dissertation fellows forum, wherein concurrent panels feature 3–4 related presentations with comments from a senior scholar who has worked with the dissertation fellows to prepare. An important recent addition to this arrangement is a first-year postdoctoral fellow who serves as the chair of each forum, and another postdoctoral fellow who serves as junior discussant, both having been mentored by the senior discussant. Serving as a session chair is usually a relatively straight-forward task, but being a discussant can be tricky, although the issues are not often made public and visible. Over time we learned that many fellows were uncertain about how to avoid the two common pitfalls—being positive by staying primarily at the level of general summary or even platitudes, or being critical but also taking the risk of appearing negative, harsh, or even condescending. Offering scholarly critique that is specific and attentive to both methodological and conceptual issues and is, at the same time, actually helpful to the scholar and supportive of his or her work, is definitely a learned skill with conventions varying across the disciplines. Providing scholarly critique is not something scholars automatically know how to do simply because they have earned a PhD, but this is a good example of a skill that can be supported and developed in the context of intergenerational professional learning.

The examples above, which portray single moments in the course of the professional lifespan of a scholar, illustrate intergenerational professional development and show how the learning of one generation builds on the learning of another. These moments are important, but they do not adequately capture the greater impact of intergenerational professional development. Beyond honing research skills and teaching the tasks of the

profession, the real power of the NAEd's intergenerational professional development approach is that it creates multiple and ongoing networking opportunities. Most importantly, scholars from other generations recognize and validate the work of emerging scholars.

In 2013, I organized a plenary session for the NAEd annual meeting, which focused on the Academy's professional development and fellowship programs. The session included comments from several former fellows whom I invited to talk about the impact of the Academy's professional development program. Laura Muñoz, a postdoctoral fellow in 2011, is now an associate professor of history at Texas A&M University-Corpus Christi. During the plenary, she shared these comments, which I use with her permission:

In a moment like this—which for me is an achievement I could never have imagined as a child or as an undergraduate student—I am reminded of my great-grandmother San Juana Peña Casas. She came to the U.S. in the early 1900s, she had no formal schooling and taught herself to read and write.... She made a conscious decision to educate her children. And after four generations, my cousins finally earned Master's degrees, J.D.'s and PhDs. Now that I am here, it is imperative for me to say that your intergenerational mentorship is critical, and has been essential to helping me learn the next steps of the profession....

Now, as I consider my life after tenure and after the post-doc, the National Academy constantly shapes the way I think about my future.... I know that in the back of my mind—in that subconscious zone where the imaginary ruminates—the post-doc reset my aspiration to build a career that would continue to garner the attention of the academy. In terms of intergenerational professional development and mentoring, the Academy members have been incredible.... Upon my selection as a post-doc, NAEd members reached out to me immediately.... These kinds of new alliances are especially critical to scholars, like myself, outside R-1 institutions.

These comments from Laura Muñoz speak to the remarkable power of intergenerational professional development and to the impact of validation from senior scholars and others. The Academy's professional development program offers unparalleled opportunities for young education researchers to be mentored by the top scholars in the field and to connect with an extraordinary network of individuals and research groups working on related issues.

However, Muñoz's comments also touch on one of the important challenges that the NAEd faces as it moves forward with its intergenerational professional development program. The program deliberately nurtures success by building on the experience of people who have already been extremely successful in the current system of scholarly knowledge

production, dissemination, and implementation. Identifying the young scholars who become recipients of the dissertation and postdoctoral fellowships is a critically important and extremely labor-intensive task, ably accomplished by the members of the NAEd Fellowship Selection Committees. Playing a key role in a selection process that is likely to have an extraordinary impact on the course of a young scholar's professional academic career is daunting, and the committee members take the task very seriously. One of the biggest challenges the NAEd now faces in terms of its professional development framework is to ensure that the process of identifying and nurturing the most promising young scholars does not simply reproduce the system of privileges and opportunities enjoyed by those who were trained at top research universities or does not identify only those new scholars whose research questions and methods most closely resemble the work of those who have gone before. The challenge now is to figure out how to preserve the tradition of high standards for rigorous educational research and, at the same time, to expand the opportunities beyond the traditional boundaries of the academic world and to nurture work that asks new questions, offers new methodological approaches, or forges new territory.

Preparing the Next Generation of Education Researchers: Reflections on the Role of the National Academy of Education

*Adam Gamoran*¹

Much of the good work of the National Academy of Education (NAEd) has been aimed at enhancing the quality of scholarship in education. Perhaps the most important aspect of that work has been to play a role in nurturing the next generation of outstanding researchers.

For more than 30 years, the NAEd's main activity in the professional development of emerging scholars has been its operation of the postdoctoral fellowship program created and supported by the Spencer Foundation, the nation's only private source of funding exclusively devoted to supporting research on education broadly defined. Since 2011, the NAEd has also operated the Spencer Foundation's second fellowship program, which supports dissertation research. Together, these two fellowship programs are among the most visible and prestigious sources of research support for future leaders of education research.² Through the involvement of its members (who participate as volunteers), the NAEd ensures not only that the fellows are among the most successful scholars at their career stages, but also that the fellowship experiences enhance the progress of their work.

¹ Adam Gamoran is the President of the William T. Grant Foundation. He was elected to the National Academy of Education in 2001.

² Information on the fellowship programs is posted at <http://www.spencer.org/fellowship-awards>.

Effectiveness of the NAEd/Spencer Fellowship Programs

In 2010, the Spencer Foundation invited me to examine the effectiveness of the fellowship programs and to place them in the larger context of alternative approaches to human capital development in education research (Gamoran & Bruch, 2011). Among the key findings of this examination were:

- The NAEd/Spencer Fellowship Programs occupy unique niches in the landscape of programs to develop researchers, in that they support researchers across all areas and approaches in education research. They are also distinctive in their support for international research and international scholars. If they were eliminated, then no other existing programs would meet the needs they fill.
- Quantitative studies by Hedges, Hanis, and Asch (2011) and qualitative analyses by Gamse and Conger (1997) and Gamse et al. (2001) revealed that the NAEd/Spencer Fellowship Programs are effective in enhancing the work and careers of education researchers. For example, regression discontinuity analyses showed that at both the dissertation and postdoctoral levels, the fellowships enhanced the scholarly productivity of fellows by 20 percent to 30 percent compared to finalists who did not win the awards (Hedges, Hanis, & Asch, 2011).
- The dissertation program (and possibly both programs) encourages outstanding young scholars, who may be interested in a variety of topics within their disciplinary fields, to focus on education as a field of inquiry. This conclusion was principally based on the finding of Hedges, Hanis, and Asch that, years after winning the fellowship, 46 percent of dissertation fellows were members of the American Educational Research Association (AERA), compared to 34 percent of finalists who did not win. Qualitative reports also showed how the awards helped focus the attention of fellows on education research (Gamse & Conger, 1997; Gamse et al., 2001).

The NAEd deserves much of the credit for the success of the postdoctoral program and, since 2011, for the dissertation fellowship program as well. The NAEd staff play major roles in organizing the competition, and members serve on the selection committee and as reviewers of applications. More importantly, staff and members have contributed countless hours of service to organizing mentoring activities for fellows. From one-on-one, to small-group, to formal lectures, the NAEd has provided a context for emerging scholars to engage with their peers, and with vet-

eran scholars, as they worked through the challenges of pursuing their research and managing their careers.

Recommendations for the Career Development of Future Scholars

Based on the findings of our review, we urged the Spencer Foundation to “maintain the brand and stick with success” (Gamoran & Bruch, 2011). In light of the success of the programs at meeting their aims, and the unique purposes they fill, we advised that sustaining the programs is important for maintaining and building the field of education research. At the time, the Spencer Foundation was rethinking the value of its financial support for the fellowship programs, and following our review and the analyses of Hedges, Hanis, and Asch (2011), it decided to continue to back the programs (McPherson, 2011).

In light of the importance of the mentoring and networking activities to the fellowship experience—former fellows often identified these as the most valuable components—some suggested eliminating the fellowship stipends and just providing the mentoring and networking support. This would have allowed the Spencer Foundation to support many more fellows, or to reduce its spending on the fellowship programs in favor of more funds devoted to major grants or other purposes. Although tempting, we advised against moving in this direction. Although we commonly think of the application process as a competition among potential fellows, it is also a competition among funders who seek to support the most talented young scholars. The Foundation and, by extension, the NAEd, compete to support the strongest emerging scholars and mark them as education researchers. Without the stipend, the best potential education researchers would find support elsewhere, and may not be the ones to receive mentoring and networking from the NAEd.

The NAEd/Spencer Postdoctoral Fellowship Program is not the only career development award for postdoctoral researchers in education. Others include the National Science Foundation’s (NSF’s) CAREER Award and the William T. Grant Scholars program, which is operated by the Foundation I lead. However, neither of these fellowship programs focuses on education research with the same breadth as the NAEd/Spencer Program, because the NSF program focuses on mathematics and science education and the William T. Grant Foundation’s award is for research on youth, until recently aimed at understanding their social settings and now on reducing inequality, as well as the use of research evidence in policy and practice.³ Neither is intended specifically for education research, nor

³ See <http://wtgrantfoundation.org/Grants#apply-wtgrant-scholars>.

admits the wide range of education researchers who enter the NAEd/Spencer Programs (Gamoran & Bruch, 2011).

In addition to recommending continued support of the fellowship programs, we also recommended that the Spencer Foundation consider a new, targeted, graduate training program in the “purposes and values of education,” one of the Foundation’s primary areas of interest. Though not exactly as we described, the Spencer Foundation has since launched a Center for Ethics in Education at the University of Wisconsin–Madison and the University of Illinois–Chicago, with support for graduate student preparation as a major component.⁴ We recommended an activity of this nature because it emerged as a missing piece in our scan of support for human capital development in education research.

Yet another feature of career development that is not currently served by the NAEd/Spencer Fellowships is targeted training in specific skills for education researchers. In our report to the Spencer Foundation, we provided three examples of specific skills for which training might be offered. As of yet, these suggestions have not taken root. First, we proposed a workshop on crafting exemplary dissertations in education research. I have written on this elsewhere (Gamoran, 2007), but reading a wide range of excellent dissertations has resulted in substantial wisdom within the Spencer Foundation and the National Academy of Education. A workshop with the NAEd members could help elevate the quality of dissertations and provide opportunities for networking to researchers who are not necessarily among the fellowship winners.

Our second idea also emerged from the Spencer Foundation’s focus on the purposes and values of education. A workshop targeted at developing tools for inquiry, to help researchers investigate such questions, build compelling arguments, and elevate the quality of discussion could be extremely valuable, especially at a time when the purpose of education is often taken for granted and not critically examined.

Our third suggestion was to provide targeted training in qualitative methods for education research. At the time of our scan, we found many more opportunities for training in specific quantitative skills than for qualitative methods skills. Revisiting this notion today, I believe that qualitative methods training is more plentiful, but targeted opportunities for mixed-methods training might be especially valuable, because this approach arguably is rarely done well and does not often receive concerted attention as a methodological skill.

⁴ See <http://ethicsandeducation.wceruw.org/index.html>.

Conclusion

The NAEd website declares that the organization “advances high quality education research and its use in policy formation and practice.” The development of young researchers is one of the main ways it achieves this aim, and with the strong support of the Spencer Foundation, the fellowship programs are a key part of its success in doing so. Both the NAEd and the Foundation are to be congratulated on this success. At the same time, it is worth considering whether there are more ways that the NAEd can advance this purpose.

References

- Gamoran, A. (2007). *Reflections on exemplary dissertations in education research*. Chicago, IL: The Spencer Foundation. Retrieved from <http://www.spencer.org/reflections-exemplary-dissertations-education-research>.
- Gamoran, A., & Bruch, S. (2011). Alternative models for human capital development in education research. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part III, pp. 27–56). Chicago, IL: The Spencer Foundation.
- Gamse, B., & Conger, D. (1997). *Follow-up study of the National Academy of Education/Spencer Foundation Postdoctoral Fellowship Program: Final report*. Cambridge, MA: Abt Associates.
- Gamse, B., Giancola, J., Griffin, N., Moore, D., Obeidallah, D., Brown-Lyons, M., & Cook, S. (2001). *Follow-up study of the Spencer Foundation Dissertation Fellowship Program: Final report*. Cambridge, MA: Abt Associates.
- Hedges, L., Hanis, J., & Asch, E. (2011). Statistical evaluations of the Spencer fellowship programs. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part II, pp. 9–26). Chicago, IL: The Spencer Foundation.
- McPherson, M. (2011). Introduction. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part I, pp. 2–8). Chicago, IL: The Spencer Foundation.

Reflections on the National Academy of Education on Its 50th Anniversary

*Larry V. Hedges*¹

I was very honored to be elected to the National Academy of Education (NAEd) in 1996. Shortly thereafter, I came to understand that this appointment was not only an honor, but also a responsibility that required work on my part when I was asked to serve on the NAEd/Spencer Post-doctoral Fellowship selection committee. The work of this committee was fascinating. In those days committee members received a large box of applications late in the fall, and we were asked to review them and assign them numerical scores. I have fond memories of spending winter days reading proposals for research in areas that were an intellectual stretch for me. The highpoint of the committee work was a winter meeting in San Francisco to conduct the first round of evaluation of the proposals. Not only was this a welcome break from brutal winters in Chicago, but also it was always intellectually stimulating work. This was true because the topics of the proposed research that the committee considered were even more diverse than the ones assigned to me (which were already stretching the boundaries of my expertise), and because the other committee members were, in short, brilliant. It was like attending a wide-ranging seminar on potentially promising topics in education research.

¹ Larry V. Hedges is the Board of Trustees Professor of Statistics and Education and Social Policy, Professor of Psychology, and Director of the IPR Q-Center at Northwestern University. He was elected to the National Academy of Education in 1996.

Selection of the NAEd/Spencer Postdoctoral Fellows

A major impetus for the program at the time was to draw scholars from other areas into research on education. In some areas, such as history and philosophy, the NAEd/Spencer Postdoctoral Fellowship was one of the few opportunities of its type and attracted fascinating projects I might never have seen otherwise. Sometimes the letters of support were of interest in and of themselves, as in the case of a proposal from a philosopher who included a letter of support from none other than Jacques Derrida (in French, but helpfully translated for those of us limited to English). The fact that we could actually understand the letter made several of us skeptical of its provenance, but I learned a lot from my colleagues on the committee who helped us understand the project.

Although my own education as a committee member was important to me, it was incidental to the real purpose of the winter meeting, which was to select finalist proposals that would undergo further review by the NAEd members. The actual selection of the fellows happened at a spring meeting of the committee, which was conducted during the annual meeting of the American Educational Research Association. This process was exciting, but difficult, because there were always more interesting proposals and worthy individuals than could be funded. In making the final selection, the evaluations solicited from the NAEd members in this second round of review were always interesting. Among other things, they revealed things about our colleagues we could only know through their evaluations, such as who in 5 successive years had the best student they have ever seen and who used a numerical score of 2 ("might have some merit") to indicate the best proposal he has seen in 30 years.

The selection process used was to first rank order the proposals by the total scores of all the available ratings (usually an average of eight ratings for each proposal). Someone would usually joke that if we needed to fund 25 fellows, we could just draw a line between the 25th and 26th highest average scores and be done in 15 minutes. None of us felt remotely comfortable doing that, however, so we discussed every proposal and argued its merits, being sure to not consider them in rank order to not be biased by the tyranny of the numerical ratings. Every year, when we were done at the end of a long day, I marveled at how the actual selections were remarkably similar (but not exactly identical) to what they would have been if we had simply drawn that line between the 25th and 26th highest average scores in the morning. I served on the committee for 8 years and enjoyed almost every minute of it.

Evaluation of the NAEd/Spencer Postdoctoral Fellowship Program

From 2003 to 2007, I served on the NAEd board. During that time, the board of the Spencer Foundation, which had been funding the postdoctoral program for almost two decades, asked the reasonable question of whether there was any evidence that the program resulted in good effects. At the time, the NAEd's largest project was the postdoctoral fellowship program, which funded part or all of many of the NAEd staff positions, so it was crucial to the NAEd's functioning. The notion that we might lose funding for the postdoctoral program certainly focused the attention of the NAEd board members. We discussed what to do and how we might evaluate the fellowship program. I recall that I made the mistake of sharing an idea based on my experience as a selection committee member: We could conduct this evaluation as a (fuzzy) regression discontinuity design. Of course it is dangerous to have ideas (or at least open your mouth about them). I found myself charged with carrying out an evaluation of the fellowship program using this regression discontinuity approach in early 2003. Fortunately, many smart graduate students from the University of Chicago were looking for support at the time, and I was able to recruit an exceptional one of them, Jennifer Hanis, as a project leader. She, in turn, recruited many of her colleagues from the Department of Sociology to help with the project.

The regression discontinuity design that I proposed for the evaluation was not new (Donald Campbell had used it almost 40 years earlier), but at the time it was rarely used. In the intervening decade it has become justifiably popular as an evaluation design with considerable merit in many practical situations. For those unfamiliar with the logic of this design, it is as follows. A fellowship program intends to award the fellowship to the most promising individuals. Therefore we would expect the fellowship recipients to be more successful than those who did not receive fellowships, even if the fellowship made no difference whatsoever. The trick in this design is to use the information available *prior to* awarding the fellowship (the evaluation scores given to the proposals for the fellowship) to predict how successful we could have *expected* participants (both those who received and did not receive a fellowship) to be. Recall that I mentioned above that the awards were remarkably similar to what would have occurred had we just used the evaluation scores (and skipped the deliberation). As it turned out, 85 percent to 90 percent of the decisions made by the selection committees we studied were identical to setting a cut point based on evaluation scores and awarding fellowships strictly based on the evaluation score.

I want to emphasize that the actual process was much more cognitive—the selection committees discussed and argued about every candidate. However a pretty good model of what we did was to follow

the evaluation scores, which means that those scores captured a great deal of what we might later characterize as pre-fellowship-decision information about the applicants' promise. Therefore, one way to describe the logic of the regression discontinuity design is that it tries to model how successful a non-recipient of a fellowship who very nearly received a fellowship would have been if he or she had received slightly higher evaluation scores and consequently a fellowship. The difference between how successful they were and (the counterfactual) of how successful they would have been had they received a fellowship is the treatment effect. The design gets its name because if the fellowship has an impact, then there will be a difference, a discontinuity, in the function describing the relation between evaluation score and career success. This discontinuity would occur at exactly the score that corresponds to the cut point where individuals whose scores are above the cut point receive the fellowship and those whose scores are below the cut point do not.

We decided to look at the 455 individuals who had been finalists from 1986 (the first year that the program ran at full scale) through 1998 (294 who were awarded fellowships and 161 who were not awarded fellowships). The latest year was selected because the fellowship starts the year after it is awarded, lasts for 1 to 2 years, and takes some time to produce its effects on career success. A fellowship awarded in 1998 would take place in 1999 to 2001, and then would take a few years for its effects to manifest in the fellow's career. Therefore, 2004, when we were gathering data, would be the first year that we could expect to see effects.

A study like this one presents many challenges, both conceptual and practical. One conceptual issue was how to frame the question, and another was what evidence might be relevant to that question. We finally settled on framing the question as whether the fellowship enhanced the careers of the scholars who received them. The issue was not the associational question of whether fellowship recipients had more successful careers, but the causal question of whether they had more successful careers *because* they received the fellowship. This framing leaves open the question of what it means to have a more successful career. After consultation with our colleagues in higher education, we realized that we had to recognize that there were several dimensions of career success, including research productivity, influence on the work of other scholars, ability to garner resources to support research, and recognition of accomplishment. Each of these dimensions in turn needed to be operationalized in ways that would allow for quantification. The interdisciplinary nature of the scholars competing for the NAEd postdoctoral fellowships made the process of selecting indicators more complex than it might have been in a single discipline. For example, productivity among historians is likely to be in the form of books, not journal articles, while the opposite is true for

psychologists. Different dimensions are also not equally important across disciplines. For example, obtaining research grants is essential among experimental psychologists, but far less important among historians.

Some logically promising ideas were rejected because we could not convince ourselves that they led to reliable evidence. For example, we thought that it would be better to divide the disciplines into groups that were more similar and assess, say, productivity among scholars whose fields had more similar concepts of scholarly productivity. Ultimately, our sample, which was as big as we could make it, was too small to support this kind of division. Similarly, we thought that prestige of appointments should be a meaningful indicator of career success, but all of our ideas of how to measure this eventually seemed fatally flawed in our highly interdisciplinary setting.

There were also many practical problems in this research. We decided to use only measures that we could extract from information on complete academic vitas. The first practical problem was locating individuals to participate in the study. We knew the applicants' contact information when they applied 5 to 17 years ago, but not what happened to them after that. In the era before social media, locating all these people was not a trivial matter (it took months of effort). Once we located them, our next task was to persuade them to participate and send us their vitas. The fellows were easy to persuade; after all we had given them a prestigious fellowship and a year off from teaching. The people who did not receive fellowships were another story. It took some diplomacy and often personalized communication to persuade them to send their vitas. People said things such as, "You guys didn't give me a fellowship so I don't know why I am doing this, but here is my vita." We were mindful that failure to get a high response rate among those who did not get fellowships would bias our results, so we persevered and acquired vitas from 78 percent of the finalists who were not fellows, which along with the 82 percent of fellows who sent their vitas resulted in the 80 percent response rate we set as a target.

Our analyses suggested quite positive effects for the fellowship program on essentially every dimension, but not uniformly on every *indicator* of every dimension. On the dimension of research productivity, the fellowship was estimated to increase the total number of publications by 27 percent, the number of articles by 20 percent, and the number of book chapters by greater than 40 percent. On the dimension of influence on other scholars, the fellowship was estimated to increase the number of appointments to editorial positions by 40 percent and the number of citations by 49 percent. On the dimension of garnering resources to support research, the fellowship was estimated to increase the number of federal grants by 80 percent. Not all of the effects were statistically significant,

but they all pointed in the same direction, indicating a positive effect of the fellowship. We presented the final report to the NAEd board in January 2005, and the Spencer Foundation board was sufficiently impressed that it extended funding for the program at the same meeting in which it discussed the report. A summary of the evaluation is presented in McPherson (2011).

This evaluation had a substantial impact on the thinking of other agencies about the evaluation of their own prizes or fellowships. For example, the Spencer Foundation subsequently commissioned a study of its dissertation fellowship program using similar methods, the National Institutes of Health funded an evaluation of research prizes using the regression discontinuity approach, and the Howard Hughes Medical Institute has begun a long-term evaluation of a new fellowship program that it is starting this year using the same methodology.

Determining the Academy's Appropriate Size

A few years later, the board considered whether increasing the NAEd's size was desirable. On the one hand, a larger membership would increase the number of individuals available for the NAEd's projects and arguably increase its impact on the field. On the other hand, a larger membership might be less distinguished, diluting the NAEd's prestige. I was asked to investigate the issue of the appropriate size for the NAEd. I decided that we could inform our understanding by studying the size of the National Academy of Sciences (NAS).

The challenge was to determine what NAEd size would result in the NAEd members being as distinguished in education as the NAS members were in their respective fields of science. I reasoned that if education researchers were typically as distinguished as science researchers, then the NAEd members would be as distinguished (on average) as the NAS members if the proportion of all researchers who were members of the two academies was the same. By this calculation, the proper size of the NAEd was driven by two factors: The number of education researchers and the proportion of them who "should" be the NAEd members (that is, the proportion of scientists who were NAS members). Arriving at an actual number for either factor is somewhat elusive. I used several approaches to obtain the number of researchers in education and in science: employment statistics; professional society memberships; numbers of doctorates awarded; faculty counts from the 96 schools in the Carnegie classification "Doctoral: Very High Research" (D:VHR); and faculty counts from the 65 schools within D:VHR that have NAS members. Each of these data sources yielded a proportion of researchers in the NAS, a number of education researchers, and therefore an estimate of the

appropriate size for the NAEed. The estimates derived from different data sources were somewhat different. Moreover the estimates using the same data source but based on different scientific fields represented in the NAS (e.g., physics, chemistry) were also considerably different. This calculation probably underestimates the appropriate size of the NAEed because many members are not appointed in departments or schools of education, while relatively few physicists, for example, who are NAS members have their only appointments in departments other than physics. However all of the estimates were larger than the size of the NAEed at that time, so they provided some empirical grounding for a decision about the appropriate size.

I have been very proud to be a member of the NAEed for the past 20 years, because of my respect for the institution and for the other members. I also appreciate the dedication of the members that have supported the Academy by giving their time and energy to its programs. We should all be grateful to them for their service, which makes the Academy work.

Reference

- McPherson, M. (2011). Introduction. In *Learning to work better: The Spencer Foundation's in-depth review of its signature fellowship programs* (Part I, pp. 2–8). Chicago, IL: The Spencer Foundation.

Modes of Inquiry for Educational Research

Hooking Up with Romantic Science

*Michael Cole*¹

For the past 40 years, anyone wishing to be present while I was conducting research was most likely to find me in some sort of after-school setting playing with children of various ages and their undergraduate buddies from the University of California, San Diego (UCSD). No matter the setting, a Boys and Girls Club, a housing project learning center, or a schoolroom after classes ended, I was likely to be moving around the room, kibitzing with the children and their undergraduate partners, or watching anxiously as a 6-year-old measured the ingredients to make a cake. To a newcomer, the comings and goings usually seemed somewhat chaotic and noisy. It did not look like a classroom, and it seemed impossible to measure cognitive or social development in such circumstances. In fact, it did not look like research at all! Instead it looked for all the world like John Dewey standing in the kitchen of the Lab School in Chicago creating a science curriculum.

Over the years, people have wondered at my seeming shift in careers, from experimental, mathematical learning theorist to mushy developmentalist who seems to ignore the most common rules for conducting research on learning and development. How did Cole come to this sorry pass? I call the kind of research I do a form of romantic science, and this essay explains how I came to adopt it.

To begin in the middle, I found myself at Indiana University conduct-

¹ Michael Cole is Distinguished Emeritus Professor at University of California, San Diego. He was elected to the National Academy of Education in 1984.

ing experimental studies of learning with Bill Estes. My subjects were rats and college sophomores. Developmental psychology was not yet on the American psychological radar, and the thought of studying the process of education had never occurred to me.

At that time, students pursuing PhDs in psychology were required to have an outside major and to pass two foreign languages. Luckily, there was support for graduate students interested in learning Russian and the historical links between Russian and American psychology (Pavlov being the leading figure), which made that topic of natural interest. While learning Russian and pursuing a Soviet Studies minor, I encountered an article by Alexander Luria that brought together my behaviorist background with the study of language, in particular, the acquisition of word meaning. In my eyes, Luria was studying learning; he was using clearly interpretable experimental techniques, and the results were exciting.

Luria was the hook. A postdoctoral year in Moscow set in motion the odd sequence of experiences that would lead me into a wholly different way to conduct empirical research on learning and development.

While in Moscow, I conducted my scientific experiment on semantic conditioning among patients with temporal lobe lesions. It produced only mud. Other experiments, all following psychophysical/experimental tradition, were published in Russia, a rarity at the time. However, I also followed Luria on grand rounds and observed how he interacted with individual patients. Although familiar with existing Anglo-American test methods of psychodiagnosis of brain injuries, he did not hold them in high esteem. Trained as a physician, he worked out some simple diagnostic methods concerning brain injury that were derived from his Vygotskian theoretical background. However, he did not use them in a rigid way; rather, he tailored how he carried out his diagnoses to the individual patient. To me he seemed like a magician pulling rabbits out of a hat. He geared his diagnostic procedures and rehabilitation strategy to the individual patient. At family tea before we left Moscow, I learned that once upon a time he had conducted research with peasants in far off Uzbekistan.

Not long after returning from Moscow, Pat Suppes and Jerry Bruner, who were rolling out the new mathematics, threw me into the Liberian hinterland, active passport in hand, because of my presumed knowledge of mathematics. That first experience of a rural, non-literate, subsistence culture provided a jolt. As a newly minted professional, it forced me to rethink the modifications to scientifically accepted experimentation that were needed if one were to take cultural context seriously when making claims about psychological processes. In addition, in pursuing this question, I was forced to rethink the nature of psychological experimentation

in general. I had fallen into the pit of contextualism and the sticky problem of ecological validity.

Simultaneously, I was being pushed to think more often and more deeply about the overall intellectual enterprise in which Luria had engaged. For the first time I began to take seriously the historical origins of modern schooling, as well as a concern that data suggesting a generalized cognitive advance as a consequence of school was badly over-generalized and misleading. My increasing concern with culture as an historical phenomenon then got mixed together with Luria's neuropsychology. I became engrossed in his work on rehabilitation of injured brain functions when my colleagues and I began to study clearly diverse children all of whom fit a social category called "learning disabled."

A next central event in my evolution was the program of research organized by my colleague Peg Griffin, who invented a variation of Luria's method of dual stimulation to develop after-school activities for children who were markedly failing to acquire literacy. These activities were carefully scripted "plays" in which children and adults used theoretically selected materials to work out the meanings of written paragraphs. The specifics of the activity are not important in the current context. What is important is that we were responsible for the children's welfare when they were in our hands. Our roles as objective experimenters were fundamentally subverted. Now we had to do more than make claims about zones of proximal development based on average differences between groups of children on some standardized measure. We were obligated to demonstrate what it means to create a zone of proximal development and how the process works in circumstances beyond the ordinary dyads of experimental studies.

And now romantic science enters my work in a serious way. Luria ends his autobiography with a description of two case studies. These endeavors (one with a mnemonist, one with a brain-injured engineer) were unlike his studies of Uzbeki peasant reasoning or the role of speech in development of self-control, or even of most patients he saw as a neuropsychologist in the clinic. Each case extended over many years, and in each he acted as both diagnostician and therapist. It is in the mixing of these two roles that romantic science emerges.

In my view, central to understanding the importance of Luria's approach, which Oliver Sacks referred to as "the dream of a novelist and a scientist combined," (Luria, 1987, p. xii), is the realization that this research allowed Luria to satisfy a lifelong ambition to resolve a central issue that had dogged psychology since its inception in the 19th century: how are we to reconcile natural science with the cultural nature of humans, and how are we to reconcile nomothetic laws that apply to populations of humans to individual, idiographic, lives? For Luria, romantic

science meant “both/and”: the complementarity of analytic “timeless” science that murders to dissect and a synthetic, time-bound, human biography to give it life.

These experiences help to explain my rationale for hanging out and participating in after-school settings where there are a lot of interesting things to do with children of different ages, including college students. My professional rationale for my activities is that my colleagues and I are engaged in “cultural-historical design experimentation.” Fittingly, our first such undertaking involved Ann Brown and Joe Campione who bravely withstood the chaos of invention and went on to do their own, seminal, form of design experimentation.

What perhaps distinguishes our manner of conducting experiments by design is our adoption of a “life course” approach. By analogy with Vygotsky’s and Luria’s insistence that the study of human development should encompass both periods of growth and decline during ontogeny, we believe that the entire life course of the designed activity is important to study to learn its properties, which are constantly changing, however permanent they may have seemed. This means, of course, that insofar as a designed activity turns out to be robust, one must continue to search for the sources of its continued development in every shifting ecology.

All of these experiences more or less explain how I got hooked on romantic science as a mode of research into the study of human development and that peculiar institution called “Modern Schooling.” It welcomes “evidence-based” research and allows for reflexive thinking about the shortcomings of its data. Such reflexivity comes from directly participating in some version of the designed activity so that you, the experimenter, can feel the flow of the activity you have helped to create.

Large societies are difficult to govern, and social scientists-as-technicians cannot avoid being caught up in the construction of instruments of governmentality. But as part of designing an activity for someone else’s own good, it is scientifically useful to get into the middle of things so you can feel the pains as well as the gains.

Reference

- Luria, A. R. (1987). *The man with a shattered world: The history of a brain wound*. Cambridge, MA: Harvard University Press.

Legislating the Value of Educational Research

*Margaret Eisenhart*¹

With Congressional passage of the Reading Excellence Act (REA) in 1998, the first-ever definition of high-quality, rigorous educational research—referred to as “scientifically based” reading research—was codified in U.S. law.

For the purposes of the REA, “scientifically based reading research” means the application of rigorous, systematic, and objective procedures to obtain valid knowledge relevant to reading development, reading instruction, and reading difficulties. It employs systematic, empirical methods of research, rigorous data analyses, and approval by a panel of independent experts or a peer-reviewed journal. (Sweet, 1998)

REA required that only reading programs whose effectiveness was confirmed by scientifically based research could be supported with funds from the federal government. REA spelled out two kinds of scientifically based research: (1) quantitative research consisting of causal, experimental studies and (2) qualitative research consisting of traditions of inquiry in the humanities that could also be assessed experimentally. This delineation of rigorous research on a major educational topic made experimental designs the “gold standard” for rigorous educational research and completely omitted contributions from social science.

This far-reaching action came about with limited input from research-

¹ Margaret Eisenhart is Distinguished Professor at University of Colorado Boulder. She was elected to the National Academy of Education in 2004.

ers and in the midst of charges from lawmakers and others that educational research lacked focus, rigor, and the ability to provide a solid base for improving educational practice and outcomes (Walters, Lareau, & Ranis, 2009). It was in this climate that Congressional legislation that became No Child Left Behind Act (NCLB) of 2001 and the Education Sciences Reform Act (ESRA) of 2002 were proposed, debated, and passed.

NCLB pertained to federal funding for *all* educational programs implemented in schools and restricted the definition of scientifically based research (SBR) even more. NCLB required research based on testing hypotheses with experimental or quasi-experimental designs and gave preference to random assignment (Eisenhart & Towne, 2003).

ESRA established the Institute of Education Sciences (IES), replacing the Office of Educational Research and Improvement (OERI) as the research arm of the U.S. Department of Education. Thus, ESRA's requirements for SBR would become the priority for *any* research on education that could be funded by the federal government.

As NCLB and ESRA were being developed in Congress, educational researchers began to take greater notice of this legislative activity. Many educational researchers believed that their voices and perspectives were not being heard. One response was the establishment of a National Research Council committee to address the question, "What is scientifically based research in education?" In 2002, the committee released its report, *Scientific Research in Education* (SRE) (National Research Council, 2002). In the report, the committee argued that scientifically based research should be defined by a set of principles appropriately applied to the research problem under investigation, not by a particular research method, for example, quantitative or qualitative, experimental or naturalistic. The principles included posing significant questions that can be investigated empirically; establishing links to relevant theory; developing a logical, evidence-based chain of reasoning; using methods appropriate to the questions posed; and replication and generalization across studies. The committee identified these principles after reviewing the characteristics of well-regarded research programs—both basic and applied—in natural science, social science, education, medicine, and agriculture.

Arguably influenced by SRE and other educational researchers and practitioners, the final version of ESRA described scientifically based research in somewhat broader terms than did NCLB. In ESRA, scientifically based research was defined as applying "rigorous, systematic, and objective methodology to obtain reliable and valid knowledge relevant to educational activities and programs, ... [using] systematic, empirical methods that draw on observation or experiment; ... relying on measurements or observational methods that provide reliable data; ... [and] ensuring that studies and methods are presented in sufficient detail and

clarity to allow for replication or, at a minimum, to offer the opportunity to build systematically on the findings of the research" (Title I, Section 102(18)). ESRA did not require experimental methods, and it included the key point that research methods should be "appropriate to the research being conducted," rather than imposed *a priori*.

More significant, however, was the fact that ESRA's mandate for IES created priorities narrower than the bill's language about scientifically based research might suggest. The general purpose of IES was to reform the field of education to be "evidence-based" and thus provide a suitable foundation for educational decision-making. But more specifically, IES was charged with (1) reporting on conditions of education; (2) identifying practices that support learning and improve achievement; and (3) evaluating the effectiveness of educational programs (Title I, Part A, Section 101(b)(1)). In practice, "evidence-based education," or "evidence based on scientifically valid research," as this approach has also been called, prioritized a research agenda focused on achievement (standardized test outcomes) and effectiveness ("what works" to improve achievement) and the use of experimental designs and randomized controlled trials whenever possible.

In the years since NCLB and ESRA became law, there has been much discussion in the educational research community about the definitions and applications of scientifically based research in education. Because both laws are now overdue for reauthorization, the topic is again on the table in Congress, and it is crucial for knowledgeable groups to make their voices heard on this important topic.

In 2011, the American Educational Research Association (AERA) took an important step when it released a list of recommendations for the reauthorization of ESRA and IES (American Educational Research Association, 2011). Among them was the recommendation that the IES encourage "a range of rigorous research methods" that are "scientifically based." Appended to the report was the AERA's statement of principles for scientifically based research:

- I. The term "principles of scientific research" means the use of rigorous, systematic, and objective methodologies to obtain reliable and valid knowledge. Specifically, such research requires:
 - (A) development of a logical, evidence-based chain of reasoning;
 - (B) methods appropriate to the questions posed;
 - (C) observational or experimental designs and instruments that provide reliable and generalizable findings;
 - (D) data and analysis adequate to support findings;
 - (E) explication of procedures and results clearly and in detail,

- including specification of the population to which the findings can be generalized;
- (F) adherence to professional norms of peer review;
- (G) dissemination of findings to contribute to scientific knowledge; and
- (H) access to data for reanalysis, replication, and the opportunity to build on findings. [not the entire statement; abbreviated for direct relevance to this article]

During the 2013-14 legislative session, the U.S. House of Representatives Education & The Workforce Committee summarized the context of this reauthorization, which is now titled Strengthening Education through Research Act as follows:

Since passage of the law, critics have contended the new standard is too strict.... For example, a limited number of programs have been found to meet the SBR definition, making it irrelevant in helping to increase the prevalence of research based programs in the classroom. Many programs exist that have proven effective in increasing student achievement, though they did not meet the rigor of SBR.... The bill replaces “scientifically-based research standards” with the term “scientifically-valid research” (SVR). The move to SVR upholds a strong standard of research, but is more inclusive than the current definition of the range of valid research methodologies utilized by IES. For example, the new definition allows IES to focus on those studies beyond experimental design and randomized controlled trials that help increase the relevance of the agency’s work. (Education & The Workforce Committee, n.d.)

As of May 8, 2014, the House passed the reauthorization bill that amended ESRA’s paragraph 18 to read:

- (18) **PRINCIPLES OF SCIENTIFIC RESEARCH.**— The term “principles of scientific research” means principles of research that—
 - (A) apply rigorous, systematic, and objective methodology to obtain reliable and valid knowledge relevant to education activities and programs;
 - (B) present findings and make claims that are appropriate to, and supported by, the methods that have been employed; and
 - (C) include, appropriate to the research being conducted—
 - (i) use of systematic, empirical methods that draw on observation or experiment;
 - (ii) use of data analyses that are adequate to support the general findings;
 - (iii) reliance on measurements or observational methods that provide reliable and generalizable findings;

- (iv) strong claims of causal relationships, only with research designs that eliminate plausible competing explanations for observed results, such as, but not limited to, random-assignment experiments;
- (v) presentation of studies and methods in sufficient detail and clarity to allow for replication or, at a minimum, to offer the opportunity to build systematically on the findings of the research;
- (vi) acceptance by a peer-reviewed journal or critique by a panel of independent experts through a comparably rigorous, objective, and scientific review; and
- (vii) consistency of findings across multiple studies or sites to support the generality of results and conclusions" (H.R. 4366, 113 Cong., 2014).

This bill was introduced in the U.S. Senate in January 2015 retaining the language from the House bill but has yet to be acted on as of June 28, 2015.

The language in the proposed Strengthening Education through Research Act is somewhat broader and more inclusive of a range of educational research methods than its predecessors. However, by focusing on methods and their technical aspects, this legislation, like its predecessors, directs attention away from the crucial prerequisite of any research activity: addressing important questions.

The focus of federal educational research should always be on what needs to be learned and better understood about the educational issues the country is trying to address. There is no lack of issues in this category, and questions about them, audiences for them, and implications of them are not limited to the narrow policy issue of what works to improve achievement as measured by standardized test scores. As Walters and Lareau (2009) clearly show based on analyses of citation counts, award lists, and expert surveys, high-quality educational research (i.e., educational research that *has been* impactful) falls into two broad and distinct categories: (1) research that has influenced scholarly debate in significant ways and (2) research that has influenced educational policy-making in significant ways. Research that has had a major influence on scholarly debate asks "quite broad questions about complex educational processes and outcomes" (Walters & Lareau, 2009, p. 203). It is theoretically informed and often disconnected from the country's political agenda at the time. Research that has influenced policy asks focused and more narrow questions, often without attention to theoretical issues and consistent with the political agenda of existing interest groups (Walters & Lareau, 2009, p. 211). Importantly, Walters and Lareau point out that few of these highly

influential and respected studies, whether scholarly or policy oriented, have been based solely or even mostly on empirical work, even fewer on experimental empirical work. Influential research studies do more than confirm or reject what is already expected. They identify unanticipated problems in need of attention. They draw attention to issues otherwise obscured or invisible. They include broad syntheses of theoretical ideas and empirical results on a given topic from multiple fields and methods. They formulate problems and solutions in new ways, and they challenge and critique conventional wisdom.

In my view, efforts since 1998 to legislate the value of educational research have missed the mark. Of course it is important to develop, articulate, refine, and justify methods for educational research, but methods are not what make educational research valuable or useful to the country. Educational research is valuable and useful when it addresses—in whatever ways available—the foremost questions of educational practice and leads to deeper understandings and new insights. It is these “broad questions about complex educational processes and outcomes,” not the methods of addressing them, that should be the national priority for educational research.

References

- American Educational Research Association. (2011). *Report and recommendations for the reauthorization of the Institute of Education Sciences*. Washington, DC: Author. Retrieved from <http://www.aera.net/ResearchPolicyAdvocacy/MajorInitiatives/IESReauthorization/tabid/10311/Default.aspx>.
- Education & The Workforce Committee, U.S. House of Representatives. (n.d.). *Bill summary: The Strengthening Education through Research Act*. Retrieved from http://edworkforce.house.gov/uploadedfiles/bill_summary_-_the_strengthening_education_through_research_act.pdf.
- Eisenhart, M., & Towne, L. (2003). Contestation and change in national policy on “scientifically based” research in education. *Educational Researcher*, 32(7), 31–38.
- National Research Council. (2002). *Scientific research in education*. Committee on Scientific Principles of Educational Research. R. J. Shavelson & L. Towne. (Eds.). Washington, DC: The National Academies Press.
- Sweet, R. (1998). *The Reading Excellence Act: A breakthrough for reading teacher training*. Washington, DC: National Right to Read Foundation. Retrieved from <http://www.nrrf.org/learning/the-reading-excellence-act-a-breakthrough-for-reading-teacher-training>.
- Walters, P., & Lareau, A. (2009). Educational research that matters: Influence, scientific rigor, and policymaking. In P. Walters, A. Lareau, & S. Ranis (Eds.). *Education research on trial: Policy reform and the call for scientific rigor* (pp. 197–220). New York: Routledge.
- Walters, P., Lareau, A., & Ranis, S. (Eds.). (2009). *Education research on trial: Policy reform and the call for scientific rigor*. New York: Routledge.

Evidence and Advocacy

Michael J. Feuer^{1,2}

A popular wall poster in the offices of friends in the policy analysis business warns that “the plural of anecdote is not evidence.” It is our way of reminding ourselves, and others, that we hold to high evidentiary standards when it comes to programs or policies that affect the lives of people or the workings of organizations. As slogans go, this one seems especially apt in the world of education, where the cacophony of opinion and advocacy often seems unencumbered by data or the findings of formal research. Although it has long suffered from an “[undeserved] awful reputation,” as Carl Kaestle (1993) so eloquently explained, there is good reason to wish for education policy and reform to become even more “evidence-based.”

Like many catchy and intuitively appealing phrases, though, this one is not easily translated into rules or standards, in large part because the word “evidence” has many definitions. Historians, lawyers, epidemiologists, economists, political scientists, biologists, teachers, mathematicians, journalists, statisticians, ethnographers, physicists, sociologists, psychologists, engineers, and philosophers all have their own ways to make meaning from data; and while principles of validity and reliability (commonly

¹ Michael J. Feuer is Dean and Professor of the Graduate School of Education and Human Development, The George Washington University. He was elected to the National Academy of Education in 2003.

² I am indebted to Dick Atkinson, Amy Berman, and Ken Prewitt for comments on an earlier draft and for their thoughtful suggestions.

associated with test theory) are the shared core of all evidentiary systems, the techniques used to reach defensible inferences vary.

In education, which relies on (and enriches) many disciplinary traditions, the appeal for evidence leads inevitably to arguments about metrics, measurement, and research designs. Efforts to reduce those arguments to a one-method-fits-all mentality, to lock scientific research into a particular evidentiary schema, are foolish. Not to pick on one such schema unfairly, but progress in understanding planetary cycles, evolution, or for that matter the effects of smoking on lung cancer has not relied on randomized trials, so it is not clear why we in education should assume that is the best or only design relevant to our issues. We need to embrace a sensibly catholic attitude toward evidence, while holding steady in our zeal for empirical and logical rigor in the framing of inferences and claims. The “hunt for causes,” as the philosopher Nancy Cartwright (2007) has argued, requires a diversity of method.

Further complicating matters is the role of personal experience, more salient in education than, say, astronomy, evolution, or the laws of relativity. As even some of our most pluralist-minded believers in democratic education can occasionally be heard to groan, everyone who has been to school—and that is just about everyone—thinks he or she is qualified to render judgments about schools and schooling. This contributes to “the awful reputation” and fuels the unfortunate and unfounded suspicion that education is “squishy”—that we make decisions by rolling the dice or checking our horoscopes. But there is context here.

For one thing, having devised an intentionally fragmented system of education that privileges argumentation, broad participation, diversity, distributed governance, individual creativity, and locally inspired innovation over conformist centralism, why should we be surprised—and annoyed—by the intrusion of values and personal memories in advocacy for reform? It seems to be a price we pay, willingly if somewhat grudgingly, for the benefits of mass public schooling that has always been more of a grass roots operation than in most other countries—and has, until recently at least, educated a greater share of the population than in most other countries as well (e.g., Goldin & Katz, 2008). The “fetish for local control,” as Richard Elmore (2000) once quipped during a National Research Council meeting, extends his argument that our “schools are ... almost always aboil with some kind of ‘change’” (p. 7)—little of which could be called “evidence-based.” Not that local control is always a good thing, but still the question is how to respect individual experience without letting it stifle objective analysis or delay social justice.

Of course, the propensity to disqualify personal opinion is often a function of the disqualifier’s own personal opinions. This brings up the thorny problem of ideology: predispositions, preferences, political lean-

ings, “priors” that “creep in at the ground level, in the precognitive state” (Schumpeter, 1954)—these things matter. Now, the ways in which values, experience, tastes, and ideas add up are complicated, but my point is that we live with these tensions because we choose to. Other societies have tried, usually with disastrous consequences, to suppress values and individual choice (Koestler, 1961) or, going to the other extreme, to suppress science when it interferes with politics or religion (Ibsen, 1882). Rather than dismiss ideology as a pollutant in the pristine waters of science, we should figure out ways to account for it and, if we care about democracy, celebrate it even while acknowledging and dealing with the biases it may introduce.

There is more. It is not just that personal experience and opinion matter because we like inclusive democracy and see education as pursuit of “the good life” (Cremin, 1991). It is also because professional practice is a vital source of knowledge needed to inform theory and design useable research. Experiential evidence, what Lee Shulman (2004) memorably called “the wisdom of practice,” must be an accepted ingredient in the recipe for research oriented to use. Why shouldn’t the experiences of classroom teachers be part of the foundation for rigorous theory about the improvement of teaching? Why shouldn’t the complexities of university administration inform theories of institutional change?

In other words, it is not only an ethical stance, that is, to include the insights and values of working professionals and to take account of the effects or consequences of policy on people and organizations (Messick, 1995). It is also an *academic* imperative, in the sense that practical knowledge is essential to the construction of useful theory. Along these lines, I have argued elsewhere that education researchers can learn from “institutional economics,” especially as revived by Jim March, Herbert Simon, and Oliver Williamson, who enriched economic and organization theory with realities of human decision-making (March & Simon, 1958; Williamson, 1975). The bridge between research and practice has two lanes: travelers should pass with care.

Switching metaphors of risk from highway to circus, the tightrope we tread requires constant attention to balance. Education is fundamentally political (Henig, 2008; Jennings, 2015), which is a good thing; the downside is when partisanship impedes progress. One way to live with the tension is to consider the alternatives: most of us would not want to return to pre-enlightenment rejection of factual evidence, nor would we want to deny the validity of experience and personal values. Ken Prewitt (National Research Council, 2012) has offered a wise compromise, a distinction between “evidence-based” policy and “evidence-influenced” policy, the latter to suggest that science can and should be an important input to decisions that are ultimately politically driven and value-laden.

Managing in this environment, finding a reasonable middle ground for objective evidence to inform something as political and emotional and personal as education, requires more than clever posters or exhortations for “rigor.” It is not a task for the impatient. We invite and pay for an extraordinary amount of certifiably expert input to feed our apparently insatiable appetite for data. (I know because through much of my career I worked in organizations that try to satisfy this hunger.) But how all those data are used, what funders and consumers of the data actually believe they are getting, and what happens to the information in the long slow grind of policymaking, remain murky (the fine work of Carol Weiss [1977], David Cohen and Charles Lindblom [1979], Lorraine McDonnell [2009], and others notwithstanding).

Among the innovations to connect science and politics and prevent creeping ideology from running off in a full gallop, the founding of the National Academy of Sciences (NAS) in 1863 (Feuer & Maranto, 2010) and the National Academy of Education (NAEd) 102 years later stand out. They are exemplars of an American (and some cynics might argue, naïve) desire to bring empirical evidence and expertise from many fields and disciplines into deliberations that would otherwise be controlled entirely by politics or—God forbid—religion. The Academies share a number of values, challenges, and procedures, albeit under different legal and institutional arrangements (the NAS has a congressional charter and its operating arm, the National Research Council [NRC], was established by Executive Order). Herewith three lessons I have learned from my time at the NRC and my involvement with the NAEd, which I hope may help us prepare for a rich and productive “second-fifty.”

1. *Evidence as a cause.* A fine NRC staff member told me, during her exit interview, that after a while it was just too frustrating to not be able to advocate publicly for the causes she believed in. It was a good lesson. The credibility of the institution, in this case the NRC but equally applicable to our NAEd, hinges in part on “optics”: we all have our priors, so the question is how to reduce the risk that good scientific evidence will lose its value only because of researchers’ personal viewpoints and their desire to advocate for change. In a world crowded with so-called think tanks and policy shops that peddle partisanship masquerading as research, what Alice Rivlin (1973) once called “forensic social science,” one advantage of the Academy depends on keeping evidence ahead of advocacy—even if we are not sure how to define evidence and appreciate the passions that bring people to this work in the first place.

2. *Procedural rationality.* Most of the problems addressed by the NAS and the NAEd do not have optimal solutions; they are at best amenable to evidence-informed guidance toward what might be called “best-bet” strategies. To borrow with awe and gratitude from Herbert Simon (1976),

I believe that rational policy analysis is the search for reasonably good solutions based on appropriate deliberation, rather than the search for definitive solutions based on exhaustive computation (this was the main theme of my 2006 book). Process matters, therefore, but herein lies another paradox: the credibility of advice hinges on the rigor of the methodologies used to generate it, but those rigors are expensive and time-consuming—sometimes so much so that when the answer finally appears the original question has been forgotten or replaced. Setting the dial of “appropriateness” means looking for a balance between standards of inquiry and timely relevance.

Some examples will be familiar. The NRC consensus process is a unique approach to producing knowledge via interdisciplinary synthesis of data; but it can be a tad frustrating for obvious reasons. (Catherine Snow, an experienced committee member and chair [e.g., National Research Council, 1998], said it reminded her of Dilbert, the cartoon character who remarked on the folly of writing a sentence by committee.) Ditto for the NRC report review, which is often the butt of humor because it appears to privilege rigor over relevance; findings may not be exhaustive, but the process can be exhausting. Still, casual suggestions for efficiency need to be weighed against the threats to credibility. Maybe one day there will be an NAS report called *The Timeliness of Science Meets the Science of Timeliness*. Meanwhile, we need to be sensitive to these issues and work on improved communications between researchers and policymakers.

3. *Complexity and its discontents.* A Nobel laureate in physics once told me he became interested in K–12 education, but after a few years of working in schools he decided searching for the origins of the universe was easier. (Maybe because he was not required to conduct a randomized trial?) Yes, our problems are special, although as Rich Shavelson (National Research Council, 2002) and others have noted, complexity is not an excuse for abandoning empirical methodology or lowering our evidentiary standards. However, we do need to set reasonable expectations: for how quickly we can generate valid and reliable findings especially on the most difficult questions, and for how soon those findings might find their way into policy and practice. Our capacity for delayed gratification is challenged in an era of instant messaging and 2-year Congressional cycles. There is no algorithm that will get us to some chimerical optimum; it is just something we need to always be sensitive to.

To conclude, a bit of personal indulgence. When I received notice of my election to the NAEd, in 2003, I was sitting in my NRC office, probably contemplating the report review crisis du jour, planning my next phone date with my beloved board chair and mentor, Dick Atkinson, or scheming about how to twist Bob Linn’s or Rich Shavelson’s or Catherine Snow’s or Carl Kaestle’s or Jim Pellegrino’s or Lorraine McDonnell’s or

Bob Hauser's or Ellen Lagemann's or Paul Holland's or Chris Edley's or Laurie Wise's arm into chairing yet another board or committee. (The list of spectacularly generous experts whose sabbaticals I routinely interrupted is longer, and my affection for them all is deep and permanent.) It was a moment of intense joy, coupled with plenty of disbelief. My staff colleagues may remember hearing me gasp and seeing me dance a jig in the corridors of the Keck building. Nel Noddings, then our president, became a new friend and much admired colleague. Mike Smith, David Berliner, and other pals called me with telephonic hugs. I kept my letter in a safe file. Being elected was about the best honor I have received, one that I took as an invitation to become active in research, mentorship, and other things we do. Here is a toast to our community, and a loud cheer for our future.

References

- Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge, UK: Cambridge University Press.
- Cohen, D., & Lindblom, C. (1979). *Usable knowledge: Social science and social problem solving*. New Haven, CT: Yale University Press.
- Cremin, L. (1991). *Popular education and its discontents*. New York: Norton.
- Elmore, R. (2000, Winter). *Building a new structure for school leadership*. Washington, DC: The Albert Shanker Institute.
- Feuer, M. (2006). *Moderating the debate: Rationality and the promise of American education*, Cambridge, MA: Harvard Education Press.
- Feuer, M., & Maranto, C. (2010). Science advice as procedural rationality: Reflections on the National Research Council, *Minerva*, 48, 259–275.
- Goldin, C., & Katz, L. (2008). *The race between education and technology*. Cambridge, MA: Belknap Press for Harvard University Press.
- Henig, J. (2008). *Spin cycle: How research is used in policy debates, the case of charter schools*. New York: Russell Sage Foundation/The Century Foundation.
- Ibsen, H. (1882). *An enemy of the people*.
- Jennings, J. (2015). *Presidents, Congress, and the public schools: The politics of education reform*. Cambridge, MA: Harvard Education Press.
- Kaestle, C. (1993, Jan.–Feb.). The awful reputation of education research. *Educational Researcher*, 22(1), 23+26–31.
- Koestler, A. (1961). *Darkness at noon*. New York: Scribner.
- March, J., & Simon, H. (1958). *Organizations*. New York: Wiley.
- McDonnell, L. (2009, August). Repositioning politics in education's circle of knowledge. *Educational Researcher*, 38, 417–427.
- Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741–749.
- National Research Council. (1998). *Preventing reading difficulties in young children*. C. E. Snow, M. S. Burns, & P. Griffin (Eds.). Washington, DC: National Academy Press.
- National Research Council. (2002). *Scientific research in education*. R. J. Shavelson & L. Towne (Eds.), Committee on Scientific Principles for Education Research, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

- National Research Council. (2012). *Using science as evidence in public policy*. K. Prewitt, T. A. Schwandt, & M. L. Straf (Eds.), Committee on the Use of Social Science Knowledge in Public Policy, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Rivlin, A. (1973, April). Forensic social science. *Harvard Educational Review*, 43(1), 61–75.
- Schumpeter, J. (1954). *History of economic analysis*. New York: Taylor and Francis.
- Shulman, L. (2004). *The wisdom of practice: Essays on teaching, learning, and learning to teach*. S. Wilson (Ed.). New York: Jossey-Bass.
- Simon, H. (1976). From substantive to procedural rationality. *Method and appraisal in economics*, S. Latsis (Ed.). Cambridge, UK: Cambridge University Press.
- Weiss, C. (Ed.). (1977). *Using social research in public policy making*. Lexington, MA: Lexington Books.
- Williamson, O. (1975). *Markets and hierarchies*. New York: Free Press.

Office of Educational Research, Institute for Education Sciences, and Shaping Education Research

*Kenji Hakuta*¹

I write this reflective essay as an academic with experience around policies regulating the focus, quality, and utility of the federal research agency in education—the National Institute of Education, the Office of Educational Research and Improvement (OERI), and the Institute for Education Sciences (IES) across different congressional authorizations of the U.S. Department of Education. My particular perspective is a narrow one, covering the period from 1995 to 2002 when I served as chair of the OERI policy board (with the awkward but descriptively accurate acronym NERPPB—National Educational Research Policy and Priorities Board) as these issues shifted into the IES authorization through the Education Sciences Reform Act (ESRA) of 2002.

What I remember most strikingly during the early days of NERPPB, as we communicated with Congress, was the tepid nature of the enthusiasm toward research in education exhibited by members and staff. Equally amazing to me was the bipartisan nature of this sentiment. The voice of the occasional champion—Representative Major Owens of New York being one—was drowned out by the chorus of groans about how Congress only heard from the producers of the research (read: the American Educational Research Association and the interest groups representing the educational research labs and centers). By contrast, we were lectured, authorizations and appropriations for health are characterized by vocal

¹ Kenji Hakuta is the Lee Jacks Professor of Education at Stanford University. He was elected to the National Academy of Education in 1995.

representation of its consumers. Why, in education research, do we not hear from parents or community groups, or even from teachers? Besides, we were told, the research is not very good anyway. Unlike the high-quality science funded by the National Institutes of Health, what OERI funds is not really seen as science. So there I witnessed first-hand what Carl Kaestle in 1993 labeled “the awful reputation of education research” (Kaestle, 1993).

In fairness, the reauthorization of OERI that took place in 1994 as part of the Educate America Act: Goals 2000 legislation attempted to address some of these issues, among other things by establishing an arms-length distant board (NERPPB) with significant independence and authority over setting the priorities and peer-review standards for the agency, as well as its own budget. The intent was to address the focus, quality, and utility of the research through these mechanisms and standards. From early in its existence, NERPPB took this charge seriously and worked with Emerson Elliott (former commissioner of the National Center for Education Statistics [NCES]) to develop a policy statement for the board to address these issues. This document appears to have survived as an ERIC Clearinghouse document and is buried deep somewhere in the archives of the U.S. Department of Education. During the life of its board, NERPPB worked with various organizations, including the academies (the National Academy of Education and the National Academy of Sciences) to lend their expertise and credibility to the work, all the while balancing the constituencies for educational research not well represented in academia, particularly the regional labs.

The board’s views around focus and quality can be found in various places in the policy statement. Around quality, the board was very clear, having already been alerted to the clarion calls for randomized control trials as the “gold standard” for evidence, as it wrote:

The power of science comes from a combination of strong theory and data that bear on the theory. This implies endorsement of explicit ideas and agreed-upon methods for exploring and testing these ideas based on observation that has internal and external consistency. Experiments, as a classification of research, should not be scattershot or universal. Rather, they should be justified by a cumulative record of rigorous naturalistic observation and piloting. This requires knowledge of context in addition to adherence to scientific canons. While experiments in education may not be used as frequently as they should as a preferred means for investigation—for a variety of reasons, perhaps, but availability of funds is surely one such reason—“science” should not be equated with “experiments.” (National Educational Research Policy and Priorities Board, 1999, p. 4)

This view contrasted with statements from perspectives such as those held by William Bennett, Chester Finn, and Diane Ravitch, who were calling for OERI's work to be conducted by "a new, independent Education Audit Agency, dedicated to the canons of scientific inquiry and the pursuit of truth, without fear or favor" that would "strive for scientific rigor, including, to the maximum degree possible, randomized field trials" (Hakuta, personal notes). Hans Meeder and Doug Carnine took it one step further and called for a minimum N-size of 12 per condition in the experiments. These were moments approximating parody. I remember an eager Congressional staffer who had accumulated a stack of statistical and research methods textbooks, drafting legislative language that would dictate randomization procedures, sample sizes, and power analysis.

For developing priorities, NERPPB (especially through the urging of Ed Gordon who was another NAEd member serving on the board) turned to the NAEd to commission a study to help produce recommendations. The NAEd convened a panel chaired by Jim Greeno and Ann Brown. While NERPPB was extremely impressed by the academic stature of the chairs and committee members as well as with the weightiness of the report, they were perplexed by the many complexities that academics would point to for prioritizing education research. Practitioner members of the board in particular were unsure because what they wanted was to be told something like "focus on elementary literacy, algebra in middle school, and high school student engagement because there are breakthrough research opportunities."

The resulting report (National Academy of Education, 1999) did not deliver such, but it did deliver a very useful outcome, which was to introduce the field to the idea of "Pasteur's Quadrant"—at that time an obscure reference. The book by Donald Stokes (1997) argued for limiting the traditional conception of the continuum from basic to applied research and highlighted the various motivations of research. For education, the real value was to help researchers understand that they are in the research space of Louis Pasteur (use-inspired basic research), not that of Niels Bohr (low-use, basic research). Introducing Pasteur's Quadrant into the educational terminology (which I believe is largely due to Lauren Resnick who served on the committee) has been of significant value, especially with the passage of time.

I believe that the education research community at that time was being boxed into a corner—portrayed as what an (anonymous) astute observer termed "a right-wing caricature of a left-wing nut." Highly effective during those days was Reid Lyon, chief of the Child Development and Behavior Branch at the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), who had a knack for quotable nuggets echoed by critics of OERI and advocates for the "ran-

domize at all cost” position. In one congressional testimony to the House Science Committee’s Subcommittee on Basic Research (October 26, 1999), he pointed to “anti-scientific ideologies and philosophical positions have been expressed within a culture of post-modern thinking where a major premise is that there is no genuine scientific method, but rather a sense that anything and everything goes” (Hakuta, personal notes).

With the balance shifting toward the view of science as represented by randomized control trial experiments, the NERPPB strategy (crafted through discussions with Michael Feuer who at that time was director of the Center for Education at the National Academies) was to commission a study with the National Research Council to develop recommendations on the design principles of scientific research in education. This report, chaired by Rich Shavelson, was published in 2002. “The design of a study does not make the study scientific.” It went on:

A wide variety of legitimate scientific designs are available for education research. They range from randomized experiments of voucher programs to in-depth ethnographic case studies of teachers to neurocognitive investigations of number learning using positive emission tomography brain imaging. To be scientific, the design must allow direct, empirical investigation of an important question, account for the context in which the study is carried out, align with a conceptual framework, reflect careful and thorough reasoning, and disclose results to encourage debate in the scientific community. (National Research Council, 2002, p. 6)

The report, in my opinion, was only read selectively during Congressional deliberations as it reauthorized the federal authority for education research, creating the IES in 2002 through the ESRA, but its voice did serve to temper zealotry for equating science with method. That said, the tide shifted significantly with the establishment of the What Works Clearinghouse, the setting of incentives in the GPRA (Government Performance and Results Act) goals around the numbers and percentages of studies funded that used randomized control treatment (RCT), and changes in the agency culture particularly by seeking more PhDs and researchers to fill senior staff positions and more experimentalists to serve on peer-review panels.

Fifteen years since that swing of the pendulum, we can look back at the craning for scientism in educational research during that era (a topic deserving of a book) and ask whether it improved the focus and quality of research. In broad strokes, I would venture to say that the policy shift has definitely created many large changes in the form and character of educational research—this can be seen, for example, in the significant increases in the number of studies meeting the What Works Clearinghouse standards. I suspect that it has also changed the funding culture of schools

of education receiving research funds, creating a cohort of well-funded students who work in areas that attempt to achieve the gold standard of evidence. The main question that will need to be asked is whether the increased attention to rigor has also translated into increased relevance of the research to educational practice, in addition to better approaches to rigor. In terms of Pasteur's Quadrant, the question is whether the ability of the field to consider the research question's utility in practical situations has improved, or whether we just blindly pursued a form of quality that turned out to distract us from a focus on the utility of research.

References

- Kaestle, C. (1993). The awful reputation of education research. *Educational Researcher*, 22(1), 23+26-31.
- National Academy of Education. (1999, March). *Recommendations regarding research priorities: An advisory Report to the National Educational Research Policy and Priorities Board*. Washington, DC: Author. Retrieved from http://www.naeducation.org/cs/groups/naedsite/documents/webpage/naed_080864.pdf.
- National Educational Research Policy and Priorities Board. (1999). *Investing in learning: A policy statement with recommendations on research in education by the National Educational Research Policy and Priorities Board*. Retrieved from ERIC Clearinghouse ED431036 at <http://files.eric.ed.gov/fulltext/ED431036.pdf>.
- National Research Council. (2002). *Scientific research in education*. R. J. Shavelson & L. Towne (Eds.), Committee on Scientific Principles for Education Research, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Stokes, D. (1997). *Pasteur's Quadrant: Basic science and technological innovation*. Washington, DC: Brookings Institution.

The Future of Quantitative Inquiry in Education: Challenges and Opportunities

David Kaplan^{1,2}

In 2015, after a 1-year grace period, the journal *Basic and Applied Social Psychology* (BASP) issued an editorial policy banning null hypothesis significance testing (NHST). The policy now allows author(s) to submit papers that present results using NHST Procedures (NHSTP), but prior to publication “authors will have to remove all vestiges of the NHSTP (*p*-values, *t*-values, *F*-values, statements about significant differences or lack thereof, and so on)” (Trafimow & Marks, 2015, p. 1).

The BASP editorial is set up in a question-and-answer format. In response to the question of whether other inferential procedures such as Bayesian methods would be acceptable, Trafimow and Marks (2015) write, “with respect to Bayesian procedures, we reserve the right to make case-by-case judgments, and thus Bayesian procedures are neither required nor banned from BASP” (p. 1). Although it is not clear what criteria the BASP editors would use to judge the acceptability of a Bayesian analysis, the Bayesian paradigm does offer a clear alternative to NHST, which will be further articulated below.

The motivation for BASP to ban NHST rests on the editors’ view that NHST has dulled creative thinking and has done great harm to the advancement of the social and behavioral sciences. To quote Trafimow

¹ David Kaplan is the Patricia Busk Professor of Quantitative Methods at the Department of Educational Psychology, University of Wisconsin–Madison. He was elected to the National Academy of Education in 2015.

² I am grateful to Daniel Bolt and Peter Steiner for valuable discussion on this topic. The opinions expressed are mine alone.

and Marks (2015), "We hope and anticipate that banning the NHSTP will have the effect of increasing the quality of submitted manuscripts by liberating authors from the stultified structure of NHSTP thinking thereby eliminating an important obstacle to creative thinking" (p. 2). They go on to say, "The NHSTP has dominated psychology for decades; we hope that by instituting the first NHSTP ban, we demonstrate that psychology does not need the crutch of the NHSTP, and that other journals follow suit" (p. 2).

Criticisms about NHST have existed for almost as long as the paradigm itself. Early concerns about NHST were expressed by Jeffreys (1961). More recent criticisms can be found in, for example, Cohen (1994), Gigerenzer, Krauss, and Vitouch (2004), Wagenmakers (2007) and the volume by Harlow, Mulaik, and Steiger (1997), among many others. Nevertheless, the announcement of the ban by BASP was met with considerable discussion on various quantitative methods blogs as well as editorials from quantitative methodology journals. If the ban instituted by Trafimow and Marks is successful and other journals follow suit as they hope, then what impact might the ban have for the future of quantitative inquiry in education?

This paper examines the motivations for the ban by first briefly describing the general problems associated with NHST. Next, we focus on the precise interpretation of the p -value. We then examine the Bayesian inferential paradigm as a constructive way forward for quantitative inquiry in education.

Problems with NHST

A critically important component of quantitative inquiry in education is inference and model building. Whether interest centers on inferring the causal effect of a school-based intervention via clustered randomized designs or building complex predictive models using data from international large-scale educational assessments such as the Programme for International Student Assessment (PISA) or Trends in International Mathematics and Science Study (TIMSS), the goal, arguably, is to quantify whether the data provide evidence in support of our research hypotheses. To this end, a considerable amount of time is spent in introductory statistics courses laying the foundations of hypothesis testing, starting with Fisher (1971/1935) and culminating in the Neyman and Pearson (1928) framework.

An interesting aspect of NHST is that students (as well as many seasoned researchers) appear to have a very difficult time grasping its principles. Gigerenzer, Krauss, and Vitouch (2004) argued that much of the difficulty in grasping conventional hypothesis testing lies in the conflation

of the approaches advocated by Fisher (1971/1935) and by Neyman and Pearson (1928).

Briefly, Fisher's approach to hypothesis testing requires specifying only the null hypothesis. For Fisher, the term "null" meant, the hypothesis to be "nullified;" however in education research, the null hypothesis is virtually always taken to be the absence of the effect of interest.³ Next, a conventional significance level is chosen (almost always the 5 percent level). Once the test is conducted, the result is either significant ($p < 0.05$) or it is not ($p \geq 0.05$). If the resulting test is significant, then the null hypothesis is rejected. However, if the resulting test is not significant, then no conclusion can be drawn.

In contrast to Fisher's ideas, the approach advocated by Neyman and Pearson requires that two hypotheses be specified—the *null* and *alternative* hypothesis. By specifying two hypotheses, one can compute a desired tradeoff between two types of errors: *Type I errors* (the probability of rejecting the null when it is true) and *Type II errors* (the probability of not rejecting the null when it is false). The Neyman and Pearson approach is, in fact, a decision-theoretic framework, providing information leading to an action taken by the researcher. Under the Neyman and Pearson approach, studies are ideally designed prior to data collection so as to minimize Type I or Type II errors, depending on the goals of the research. The important point here is that the Neyman and Pearson approach is not a framework for quantifying evidence. Rather, it is an approach for minimizing errors.

The conflation of the Fisher approach and the Neyman and Pearson approach to hypothesis testing lies in the use and interpretation of the p -value. In Fisher's paradigm, the p -value is a matter of convention with the resulting outcome being based on the data. In the Neyman and Pearson paradigm the Type I and Type II error probabilities are determined prior to the experiment being conducted and refer to a consideration of the cost of making one or the other error.⁴ However, even a casual perusal of the top journals in education and the social sciences generally will reveal that this balance is nearly always ignored and a Type I error probability of 0.05 is used, that value itself being the result of Fisher's experience with small agricultural experiments. To quote Gigerenzer, Krauss, and Vitouch (2004): "For Fisher, the exact level of significance is a property of the data, that is, a relation between a body of data and a theory; for Neyman and

³ For this paper, an "effect" can be taken to be a treatment effect in, for example, a school-based experiment or a set of relationships (correlations or regressions) in an observational study.

⁴ The term "experiment" is used here to refer to any systematic data collection exercise and encompasses observational studies, quasi-experiments, as well as randomized experiments.

Pearson, α [the probability of a Type I error] is a property of the test, not of the data. Level of significance and α are not the same thing."

What Is the p -value?

The problem with NHST, therefore, seems to lie with the interpretation of the p -value. To be precise, the p -value rests on a form of argumentation referred to as *reductio ad absurdum*. This form of argumentation attempts to establish an assertion by deriving an absurdity from its denial, thus arguing that the assertion must be accepted because its rejection would be indefensible. The p -value as *reductio ad absurdum* is thus properly interpreted as *the probability of observing data at least as extreme as the data that was actually observed, computed under the assumption that the null hypothesis is true*.

Let's look at the interpretation of the p -value carefully. First, the p -value is based on data that were never observed. Specifically, the p -value is obtained by referencing the value of the observed test statistic (e.g., the t -test) based on the study at hand to hypothetically replicated data generated under the assumption that the null hypothesis is true. Second, as stated, the p -value is computed under the assumption that the null hypothesis is true. In most applications, the null hypothesis is taken to reflect no difference (or no effect, no relationship, etc.). To begin with, this hypothesis is never true in reality, and moreover, it is typically not the research question of interest. Thus, after Cohen (1994), researchers are typically testing a "nil" hypothesis that is hopefully rejected. However, it is important to emphasize that there is nothing within the NHST paradigm that requires testing a null hypothesis of no effect. In fact, any theoretically justifiable hypothesis can serve as the null hypothesis.

So, to summarize, the p -value does not quantify evidence for a hypothesis. Rather, it provides a measure of the probability of an outcome not actually observed, computed under the assumption of a null hypothesis that will likely not be true in any population. To quote Jeffreys, "This seems a remarkable procedure" (1961, p. 385).

Obviously, misunderstandings of NHST and the associated p -value are not sufficient to ban its use. However, these misunderstandings combined with the clear and well-documented bias toward publishing only statistically significant results have led authors to questionable practices, such as describing non-significant findings as "trending toward significance"—a nonsensical phrase, among many in use today, that has no basis in statistical theory. However, more seriously perhaps for the advancement of quantitative inquiry in education, the conventional p -value does not quantify evidence for a hypothesis of interest, and this, arguably, is what education researchers are after.

A Way Forward: The Bayesian Paradigm

What then are the opportunities for hypothesis testing in education? First, it should be noted that NHST could be valuable when rigorous error control is desired. For example, in the case of an educational intervention where the consequences of committing a Type I or Type II error has important high-stakes consequences for a student, NHST may be warranted. However, for practical applications of model building and evaluation, we argue that the Bayesian paradigm of statistical inference represents an internally consistent and coherent alternative to NHST—an alternative that is now readily available to education researchers because of the development of computational algorithms applicable to Bayesian analysis. Classic discussions of Bayesian statistics can be found in, for example, de Finetti (1974) and Savage (1954). An advanced text on the topic is Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin (2013). An introductory text on Bayesian statistics with applications to education research can be found in Kaplan (2014). An excellent treatment of Bayesian epistemology can be found in Howson and Urbach (2006).

Briefly, in contrast to conventional Fisherian statistics, which considers probability as synonymous with the long-run frequency of outcomes, Bayesian inference treats probability as the language for encoding uncertainty about those elements of an analysis that are unknown⁵; in particular, model parameters such as treatment effects or regression coefficients.⁶ Second, unknown parameters are assumed to be random variables described by probability distributions representing cumulative knowledge regarding what is reasonable to believe about the parameters of interest. These probability distributions are referred to as *prior distributions* that can be elicited from personal subjective belief, expert opinion, and/or prior research. Third, through Bayes' theorem, prior distributions on the model parameters are combined with the probability model of interest for the current data (e.g., a regression model) to yield updated knowledge about the unknown parameters summarized in the so-called *posterior distributions*. Finally, the focus of model evaluation in Bayesian inference is based largely on predictive quality and not on goodness-of-fit, per se. Statistical significance testing and all its "vestiges" (including NHST *p*-values) play virtually no role in Bayesian inference. Thus Bayesian inference is a framework for learning from data. To quote Jerome Cornfield (cited in Savage, 1954), "[I]t is clear that it is not possible to

⁵ In the case of probability as long-run frequency, the canonical example is the flipping of an unbiased coin. In the case of probability as encoding uncertainty, the canonical example is betting on the outcome of a game.

⁶ Bayesian inference can also treat the uncertainty in model building and selection through the method of Bayesian model averaging. This is beyond the scope of the paper.

think about learning from experience and acting on it without coming to terms with Bayes' theorem."

To be clear, Bayesian inference is not without its own internal set of controversies—the most salient of which is the choice of priors. Recall that priors can be obtained from personal belief, expert opinion, and/or prior research. Eliciting priors, especially those based on subjective belief, is fraught with difficulties. Indeed, without careful techniques of elicitation and the formal comparison among models with different priors, a researcher can certainly skew the results toward his/her prior beliefs. The problem of *elicitation* is discussed in O'Hagan et al. (2006). In response, many in the Bayesian world have sought so-called reference priors that retain the benefits of the Bayesian framework for uncertainty quantification while at the same time letting the data "speak" as much as possible. Such reference priors may be particularly useful in policy situations where subjective priors may not be appropriate. Thus, the Bayesian world can be roughly divided into "subjectivists" (e.g., de Finetti, 1974) and "objectivists" (e.g., Berger, 2006). Rich methodological research is continuing to flow from these two schools within the Bayesian paradigm, and it is important that those engaged in quantitative education inquiry become familiar with the debate (see, e.g., Kaplan, 2014, Chapter 10, for a discussion of the debates within the Bayesian paradigm situated within education research).

To summarize, Bayesian statistical inference offers a constructive and available alternative to NHST for quantitative inquiry in education. By explicitly recognizing and directly accounting for uncertainty, Bayesian statistical inference can guide education research toward evolutionary knowledge development and away from blind adherence to NHST. However, we do not agree with the editorial policy of BASP that NHST should be banned. In fact, as noted above, there may be situations in which the kind of error control that is achieved through the proper use of NHST is desired. Nevertheless, we would argue that such situations are relatively rare in education research. Rather, often the goal of quantitative inquiry in education is to assess whether a hypothesis of interest is supported by the data in hand. Addressing this goal forms the basis of Bayesian inference; recognizing and explicitly accounting for all manner of uncertainty that can enter into an inquiry.

For a Bayesian approach to quantitative inquiry in education to move forward, several steps are at necessary: (a) Bayesian epistemology should be formally taught alongside NHST in introductory statistics classes, clearly defining for students the challenges and opportunities of each approach, (b) authors must clearly warrant the choice of either NHST or Bayesian inference for their investigations, and (c) "best practices" in Bayesian inference must be developed and demonstrated and then these

best practices should be adopted by research journals and funding agencies. These steps would go a long way toward improving quantitative inquiry in education.

References

- Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, 3, 385–402.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997–1003.
- de Finetti, B. (1974). *Theory of probability* (Vols. 1 and 2). New York: John Wiley and Sons.
- Fisher, R. A. (1971/1935). *The design of experiments* (9th ed.). New York: Macmillan Publishing.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis* (3rd ed.). Boca Raton, FL: CRC Press.
- Gigerenzer, G., Krauss, & Vitouch, O. (2004). The null ritual: What you always wanted to know about significance testing but were afraid to ask. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage Publications.
- Harlow, L. L., Mulaik, S. A., & Steiger, J. H. (1997). *What if there were no significance tests?* Mahwah, NJ: Lawrence Erlbaum and Associates.
- Howson, C., & Urbach, P. (2006). *Scientific reasoning: The Bayesian approach*. Chicago, IL: Open Court.
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). New York: Oxford University Press.
- Kaplan, D. (2014). *Bayesian statistics for the social sciences*. New York: Guilford Press.
- Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 29A, Part I, 175–240.
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgments: Eliciting experts' probabilities*. West Sussex, England: Wiley.
- Savage, L. J. (1954). *The foundations of statistics*. New York: John Wiley and Sons.
- Trafimow, D., & Marks, M. (2015). Editorial. *Basic and Applied Social Psychology*, 37, 1–2.
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p -values. *Psychonomic Bulletin & Review*, 14, 779–804.

Education and Don Quixote

James G. March^{1,2}

Modern portrayals of human action are overwhelmingly in a calculative and consequentialist tradition. Consequentialist reasoning is the basis for most of modern social and behavioral science and preeminently for economics. Action is seen as choice; and choice is seen as driven by anticipations, incentives, and desires. These ideas trace their roots at least to the Greeks, owe substantial parts of their modern manifestation to the formulations of Jeremy Bentham, and derive much of their contemporary power from the geniuses of L. J. Savage and John von Neumann.

It is no surprise that educational institutions teach such a consequentialist theology as a sacred doctrine and also address their own problems of decision and strategy in the same spirit. They evaluate their alternatives in terms of expected consequences, implement strategies with expected outcomes that appear attractive, and seek to manage the actions of others by assuming they are similarly guided. Such practices honor ideas that

¹ James G. March is the Jack Steele Parker Professor of International Management, Emeritus at Stanford University. He was elected to the National Academy of Education in 1971.

² This short essay is a slightly revised version of remarks made on the occasion of March's retirement from the Stanford University faculty in 1995. A version of the essay was printed first in the *Stanford Business School Magazine* in 1996 (James G. March, "A Scholar's Quest," *Stanford Business School Magazine*, 64(4), 10–13), reprinted in the *Journal of Management Inquiry* in 2003 (James G. March, 2003, "A Scholar's Quest," *Journal of Management Inquiry*, 12, 205–207), and reprinted again in a book edited by Markus Hällgren in 2014 (James G. March, 2014, "A Scholar's Quest," in Markus Hällgren [Ed.], *Reflections on a Scientific Career: Behind the Professor's CV* [pp. 125–128]. Copenhagen, Denmark: Copenhagen Business School Press).

are of enormous importance in human development. It is inconceivable that we would abandon them.

Nevertheless, the ideas have their limitations. John Stuart Mill characterized Bentham, the patron saint of modern consequentialist thought, as having the "completeness of a limited man." In particular, Mill wrote, "Man is never recognised by [Bentham] as a being capable of ... desiring for its own sake, the conformity of his own character to his standard of excellence, without hope of good or evil from other source than his own inward consciousness" (Mill, 1962, p. 66).

Mill's comments on Bentham might as easily be applied to us. Our comfortable sense of completeness leads us, as it led Bentham, largely to exclude from our visions of human behavior a second grand tradition for understanding, motivating, and justifying action. This tradition sees action as based not on anticipations of consequences but on attempts to fulfill the obligations of personal and social identities and senses of self, particularly as those obligations and senses are informed by the ethos and practices of great human institutions. It is a tradition that speaks of self-conceptions, identities, and proper behavior, rather than expectations, incentives, and desires.

This second vision has become somewhat obscured in contemporary life, but it has a long and distinguished pedigree. It is captured classically in many major works of literature and philosophy but particularly in that great testament to the human spirit, *El Ingenioso Hidalgo Don Quixote de la Mancha*, published in 1606 and 1615 by Miguel de Cervantes Saavedra. When challenged to explain his behavior, Quixote does not justify his actions in terms of expectations of their consequences. Rather, he says, "I know who I am" ("Yo sé quien soy") (de Cervantes, 1605, Part 1, Chapter 5). Quixote seeks consistency with imperatives of the self more than with imperatives of the environment. He exhibits a sanity of identity more than a sanity of reality. He follows a logic of appropriateness more than a logic of consequences. He pursues self-respect more than self-interest.

As Quixote's misadventures illustrate quite vividly, following a sense of self has its own confusions and limitations, but it celebrates a non-consequentialist view of humanity. Great enthusiasms, commitments, and actions are tied not to hopes for great outcomes but to a willingness to embrace the arbitrary and unconditional claims of a proper life. Quixote reminds us that if we trust only when trust is warranted, love only when love is returned, learn only when learning is valuable, we abandon an essential feature of our humanness—our willingness to act in the name of a conception of ourselves regardless of its consequences.

The words are obviously a bit peculiar for this setting. But I think they have some mundane implications for those of us who claim to be

educators. Our involvements in education undoubtedly have many consequences that we value, but we also pursue and venerate knowledge and learning as a manifestation of faith in what it means to be a human being. When we recognize ourselves as sharing a human identity that is intertwined with traditions of scholarship, we are led to view education in ways that are somewhat less consequentialist than are the ways that have become familiar to contemporary discussions.

Recently, our metaphors of education have become indistinguishable from metaphors of markets. The problems of education are pictured as problems of creating educational programs (or public relations activities) that satisfy the wishes of customers and patrons rich enough to sustain them. It is a conception that yields useful insights and is not to be dismissed thoughtlessly. But it is a conception that fails to capture the fundamental nature of the educational soul.

An educational institution is only incidentally a market. It is more essentially a temple—a temple dedicated to knowledge and a human spirit of inquiry. It is a place where learning and scholarship are revered, not primarily for what they contribute to personal or social well-being but for the vision of humanity that they symbolize, sustain, and pass on. Søren Kierkegaard said that any religion that could be justified by its consequences was hardly a religion. We can say a similar thing about education and scholarship. They only become truly worthy of their names when they are embraced as arbitrary matters of faith, not as matters of usefulness. Education is a vision, not a production facility. It is a commitment, not a choice. Students are not customers; they are acolytes. Teaching is not a job; it is a sacrament. Research is not an investment; it is a testament.

And when someone says, as they certainly will and do, that all this is romantic madness, that any such foolishness requires a consequential justification, perhaps one that discovers an evolutionary advantage in traditions and faith, the proper answer is Quixote's: "For a knight errant to make himself crazy for a reason merits neither credit nor thanks. The point is to act foolishly without justification."³ The complications of confronting the ordinary realities of day-to-day life often confound such lofty sentiments, and I would not pretend that it is possible or desirable to ignore consequences altogether. But in order to sustain the temple of education, we probably need to rescue it from those deans, supporters, faculty, and students who respond to incentives and calculate consequences, and restore it to those who respond to senses of themselves and their callings, who support and pursue knowledge and learning because they represent a proper life, who read books not because they are relevant

³ "Que volverse loco un caballero andante con causa—ni grado ni gracias. El toque está en desatinar sin ocasión" (de Cervantes, 1605, Part 1, Chapter 25).

to their jobs but because they are not, who do research not in order to secure their reputations or improve the world but in order to honor scholarship, and who are committed to sustaining an institution of learning as an object of beauty and an affirmation of humanity.

I do not know whether any such thing is imaginable, much less possible. But if it is, then perhaps we can say that we, like Quixote, know who we are. And that, as my Scandinavian friends are inclined to say, would not be entirely bad.

References

- de Cervantes, M. (1605 and 1615). *El Ingenioso Hidalgo Don Quixote de la Mancha*.
Mill, J. S. (1962). *On Bentham and Coleridge*. New York: Harper & Row.

Philosophy and the National Academy of Education

D. C. Phillips^{1,2}

Although correlation does not establish causation, sometimes there are grounds to be suspicious. So, why is it that during the decades in which the National Academy of Education (NAEd) has been growing in size and in influence, my own scholarly field of philosophy of education has been withering—in size and in influence? It seems unlikely that there has been a direct causal link, but there might well have been an “outside” factor (indeed, perhaps more than one) that causally impacted the two phenomena and thus produced the correlation.

I will cut to the quick and give my own hypothesis about the matter; I will take the liberty of using a few philosophical terms here, which may make some readers uncomfortable. Over the past few decades there has been a rapidly increasing tendency, worldwide, to treat education as a means or tool, as an instrument or instrumental good, rather than being a good in itself or an intrinsic good. Intrinsic goods (honesty is a nice example, if education does not work for you) are things that are valuable in themselves and that are normatively required of us, even if they do not have a practical payoff. We all understand that being dishonest often pays huge dividends, but nevertheless we *should* be honest. Being a person who is disposed to act honestly, despite the potential loss of profit or whatever, is to be an admirable person. The same holds for education; it is better to be educated rather than not, not because of any potential

¹ D. C. Phillips is Emeritus Professor at Stanford University. He was elected to the National Academy of Education in 2003.

² I thank Bob Floden for helpful comments.

payoff—it simply is better! The struggle around the world to give women access to education nicely illustrates this. It is not simply a struggle to equip women for the world of work, it is a struggle for a *right*; women have a right to be educated, irrespective of the payoff, because education is an intrinsic good and like other such goods should be available to all.

The discussion might appear to have drifted away from philosophy of education and the NAEd, but not so! The link is this: over the past few decades the points made above about the intrinsic value of education have been overlooked, or swamped; education has more and more been seen as an instrument, and because philosophy of education is a field that—to a large extent—deals with value and conceptual issues the instrumental relevance of which is difficult (if not impossible) to discern, the tendency has been to judge it as irrelevant.

In what ways has education been seen as instrumental? The answer is easy to come by; simply cast one's mind back to the public and political rhetoric about education that has been prevalent during this period. There has been an enormous amount of discussion that has treated education as an instrument for national economic competitiveness and also as an instrument for fostering the economic success of both the middle class and the (small but powerful) class of the ultra-wealthy within individual societies. The schools have been castigated for not producing young workers who are equipped with the skills needed by industry and commerce. There has been discussion of the point that women who are educated are less likely to have large numbers of children, so education can be a tool with which to fight overpopulation. People who are educated are somewhat less likely to accept extremist and violent political and religious views, so education is also a convenient tool for achieving social stability. On the other hand, during this same period there has been a diminishing amount of discussion of what type of education is required to foster individual growth and autonomy (which is a discussion of intrinsic goods).

One illustrative example will have to suffice; drawn from the field of K–12 curricula it concerns the “Common Core State Standards.” Although some of these standards certainly can be interpreted as pointing to intrinsically good characteristics that should be developed in learners across the nation, this is not how the standards are introduced and justified. The website does not waste time on philosophical abstractions, but gets down to business; I have taken the liberty of italicizing the relevant words:

The Common Core is informed by the highest, most effective standards from states across the United States and countries around the world. The standards define the knowledge and skills students should gain throughout their K–12 education *in order to graduate high school prepared to succeed in entry-level careers, introductory academic college courses, and workforce training programs.*

The educational research agenda, both nationally in the United States and internationally, has been shaped by this instrumental attitude. International comparisons of educational attainment, comparisons of attainment between states within the United States, and even the way in which “attainment” is conceptualized; the massive effort to improve the quality of tests and the vast amount of time devoted to testing in schools (and the focus of instruction on material that can readily be tested); cost-effectiveness studies of educational programs; educational production functions; and the like—all of these are part of the instrumentalist syndrome. In addition, all of these have affected the makeup of the educational research community—those specialties that can be directed at instrumentally valuable matters flourish, while the discipline of philosophy, focusing as it does on intrinsic goods and conceptual issues, starts to wither.

My own institution is a relevant case-in-point; clearly it is a research-training institution, and the makeup of the faculty gives an indication of the views that are held about the nature of research. When I arrived in the early 1970s there were—quite unusually—two economists of education on the faculty, and there also were two philosophers of education (including yours truly); doctoral students had to satisfy several distribution requirements, one of which was to take at least two courses that dealt with normative or value issues. Nowadays the school is down to one philosopher; the normative requirement is long gone, and indeed there is hardly anyone around who can explain what the term “normative” means. Instead of spending time discussing the aims of education, the faculty spend countless hours discussing the precise nature of the quantitative research methods sequence (the fact that it is a sequence is telling—the normative requirement was never honored by being promoted to a sequence). Furthermore, economists are thriving (and, it must be stressed, they are doing excellent work, often in collaboration with economists outside the school and even outside the university).

Inevitably this climate has impacted the NAEd. I recall a meeting of the NAEd some years ago where we were asked what it was that made membership attractive to us. I remarked that it was the great diversity in disciplinary backgrounds among the members and the scholarly give-and-take that this made possible—why, I remarked in awe, some of us are economists, and others of us are ... not economists. This was treated as a nice off-the-cuff joke, but I had intended it as an insightful critical comment (although, to stress again, it is not the economists or their work that I am critical of, but the attitude of almost all of us that this kind of work sets the standard for *all* educational research).

One senses that there is a certain kind of inevitability here: The climate influences (and is influenced by) research funding priorities; this in

turn influences the research “market place”—the types and quantity of research that is pursued and that eventually gets published in the most widely read journals. The best—the “sexiest”—of this research is honored, and those who have produced it get nominated for membership in the NAEd. (Those who carry out research on what are regarded as trivial, non-instrumental, or non-cutting-edge topics generally do not get honored in this way.) The result is that, in a sense, the NAEd membership is shaped by those people, and those factors, that influence the research market place. In short, membership becomes skewed towards what the current conception is of good research on instrumental topics. The perspicuous reader will have sensed that I have been preparing the way for a somewhat radical suggestion, namely, that it is one of the functions of a learned society such as the NAEd to keep alive all scholarly research traditions, even traditions that do not contribute directly to what currently are regarded as “sexy” instrumental topics—for all these traditions have contributed to what John Dewey occasionally termed “the funded wisdom of the race.” How the NAEd can do this is a difficult matter about which my well of inspiration has run dry.

One issue on which I keep running aground (to change the figure of speech yet again) is the following: When a research tradition is undervalued—or worse, when it is regarded as completely irrelevant—there is a tendency for its practitioners to turn inward, to “circle the wagons,” to pursue esoterica, to indulge in modes of discourse that to “outsiders” in the dominant traditions seem sterile at best, and even to develop a degree of hostility toward these outsiders. Karl Popper’s words, written in a slightly different context, are relevant here: “their hostility to the society in which they live is, I think, a reflection of their unconscious dissatisfaction with the sterility of their own activities” (Popper, 1976, p. 196). It is a difficult case to make that, under these conditions, members of such a research group should be elected into membership of an organization such as the NAEd, or should receive funding from an external agency, or should be allocated precious university billets. Sadly I must confess that I regard the present situation in philosophy of education to be of this kind. There is some fine, rigorous work being done, but it is not the norm. This admission might be taken as a refutation of my suggestion that the NAEd should somehow act to keep fields such as philosophy of education alive, for who wants to foster sterility? But I think my honest confession strengthens the case for the NAEd action: the field is in danger of becoming entirely irrelevant, but this, of yet, has not quite happened. Preserving an endangered species is a noble thing, and the NAEd should accept the challenge—even though it means acting in service of an intrinsic and not an instrumental good.

Reference

Popper, K. (1976). Reason or revolution? In T. Adorno, et al., *The positivist dispute in German sociology*. New York: Harper Torchbooks.

Reflections on Scientific Research in Education

Richard J. Shavelson^{1,2}

The special task of the social scientist in each generation is to pin down the contemporary facts. Beyond that, he shares with the humanistic scholar and the artist in the effort to gain insight into contemporary relationships, and to align the culture's view of man with present realities. (Cronbach, 1975, p. 126)

In 2002 the National Research Council (NRC) published a monograph titled *Scientific Research in Education* (SRE) (National Research Council, 2002) that received widespread attention in the education research community, and ultimately influenced the definition of scientific research in education used in federal legislation. The monograph was written hurriedly, in 1 year by a committee of scholars, policy analysts, and practitio-

¹ Richard J. Shavelson is Partner and Chief Scientist, SK Partners, LLC, and Margaret Jacks Professor of Education Emeritus at Stanford University. He was elected to the National Academy of Education in 1997.

² At the time SRE was conceived, deliberated, and published, Michael Feuer headed the Center for Education at the National Research Council and was instrumental in all phases of the project. I appreciate his reading an earlier draft of the paper and his successful twisting of my arm to prepare this short paper. I am indebted to David Berliner and Gary Fenstermacher for their comments on an earlier draft; I learned a great deal from both of them and incorporated their suggestions as best as I could. Of course I take full responsibility for what is written here.

ners³ at the request of the federal National Educational Research Policy and Priorities Board (NERPP) chaired by Kenji Hakuta. The rush to get the National Academy of Sciences' (NAS's) statement of what constitutes scientifically based educational research was precipitated by recent (at the time) federal reports and proposed No Child Left Behind legislation. The federal government was getting into the business of defining what constituted scientific research in education; these definitions relied heavily on causal research methods (primarily hypothesis testing and randomized experiments) (see Eisenhart & Towne, 2003; Feuer, Towne, & Shavelson, 2002). Kenji and the Board were deeply concerned; NERPP wanted the NAS's take on what constituted scientific research in education.

Sketch of SRE's Findings and Recommendations and Reactions to Them

SRE said that education research could be and much of it is scientific, akin to other social and life sciences and even to more distal natural sciences. It said that what makes research scientific is not the method used but rather its adherence to a set of principles taken together: pose significant questions that can be investigated empirically, link research to relevant theory, use methods that permit direct investigation of the question, provide a coherent and explicit chain of reasoning, replicate and generalize across studies, and disclose research to encourage professional scrutiny and critique (National Research Council, 2002, p. 52). It also said that research design and methods should follow from the research questions posed: (a) "What's happening?"—calling for quantitative and qualitative research methods *describing* an educational phenomenon; (b) "Is there a systematic effect?"—calling for methods that establish evidence of a causal effect and giving priority to randomized experiments when feasible and appropriate but also including quasi-experiments and observational studies; and (c) "Why or how is it happening?"—calling for both qualitative and quantitative research into the *mechanism(s)* that gave rise to the causal effect.

The monograph drew considerable attention both in the policy community and the education research community. For the most part, policymakers and policy analysts found SRE to be useful; a vocal minority in the research community did not. There ensued journal issue after issue devoted to praising but mostly drawing out perceived limitations of

³ Richard J. Shavelson (Chair), Donald I. Barfield, Robert F. Boruch, Jere Confrey, Rudolph Crew, Robert L. DeHann, Margaret Eisenhart, Jack McFarlin Fletcher, Eugene E. Garcia, Norman Hackerman, Eric Hanushek, Robert Hauser, Paul W. Holland, Ellen Condliffe Lagemann, Denis C. Phillips, Carol H. Weiss with Lisa Towne (Study Director).

SRE (e.g., special issues of *Educational Researcher*, *Teachers College Record*, *Qualitative Inquiry*).

Eisenhart (2005), in responding to the criticisms, clearly and concisely laid out the contributions and limitations of SRE. Importantly, and in agreement with some of the criticisms, she called for more than traditional science to improve education research: (1) explore the implications of interpretative science in education research (human intention, intentional causation, and social interaction); (2) enhance the role that philosophy, history, ethics, and literary criticism plays; (3) recognize the role played by critical research in a free and democratic society—critics, skeptics, and multiple perspectives are needed and taken seriously; (4) recognize the “infeasibility and absurdity of experimental and quasi-experimental designs for many if not most research questions in educational research” (p. 56); (5) attend to the practical relevance of research not just its internal validity; and (6) seek agreement on criteria for high-quality education research, be it quantitative or qualitative.

Two Brief Stories

In what follows, I relate two stories that have influenced my thinking about what was and was not accomplished with SRE. I believe at the time we took a contextually appropriate stand where Congress and many in the country doubted the rigor, validity, or utility of education research, one that was balanced with respect to the committee’s charge: to define what is meant by scientific research in education. A question that we were not charged with answering and one we did not answer was: “Is scientifically based research the only or the best approach to meaningful studies of educational phenomena?” (Eisenhart & Towne, 2003, p. 31). Much of the criticism leveled at SRE was focused on the question we did not address and consequently this question is at the forefront of my reflections.

Story 1: Meeting and Reflections on Meeting with Lee Cronbach

I recall a fall 2000 meeting with Lee Cronbach in a park in Palo Alto, California. I told him I had agreed to chair the NRC committee and that the product of our work would be the next in the series that included his important work with Pat Suppes: *Research for Tomorrow’s Schools: Disciplined Inquiry for Education*. I am not sure whether Lee thought that this was something I should do. About all he said was to be careful not to oversell randomized trials in spite of the pressure to do so at the time. For him, a randomized experiment, even if done well and appropriately for the context, was just a case study of a particular situation or situations;

he doubted their ability to generalize, as noted in the introductory quote. For him, interactions of persons, treatments, and contexts were complex and abounded (Berliner & Glass, in press) much as an image multiplies in a hall of mirrors. Simply put, social and behavioral sciences are filled with interactions and unanticipated consequences (e.g., who would have anticipated that recruiting African American faculty to Stanford and Harvard would have negatively impacted Historically Black Colleges and Universities?).

I have reflected on his viewpoint over and over again for the past 40 years, asking myself, "How do I know?" At one time I had rejected his position as too soft, one that opened the flood gate for anything goes in education research. The question for me was: If I cannot control extraneous influences in reaching an interpretation and making a policy or practice recommendation, how do I know? Maybe, I surmised, Donald Campbell and his colleagues were right—randomized trials address the question of how do I know by ruling out many counter-interpretations. Such trials have been called the "gold standard." But Cronbach was also right. Generalizations decay in social research—perhaps due partially to new contexts in which an educational "treatment" is embedded (e.g., Shavelson & Webb, 2009) and to Campbell's law: "The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." I was caught in between Cronbach and Campbell and probably will continue to be. But for now I find comfort in what Cronbach said in the quote—the social scientist can at best pin down current facts as we understand them, and randomized experiments if appropriate and well conducted can help in that endeavor—no more and no less.

I believe this position is consistent with post-positivism (Phillips & Burbules, 2000), a view of knowing that endorses interactions, situations, contexts, human will, intention, intelligence, and the like. Perhaps one of the major failings of SRE was in not dedicating a chapter on this philosophical stance and showing how scientific research in education might address the complexities of the human world.

Story 2: When Scientific Research Meets Practice

While serving as dean at Stanford, the superintendent of the Palo Alto schools called and asked if I and the Stanford faculty might help his district address a pressing and heated problem: Whether or not to offer full-day kindergarten. For me this seemed like a no brainer: provide full-day kindergarten! Given the rising diversity of students in the district there was a pressing need. However, given the high proportion of Stan-

ford faculty and spouses' children in the schools, the spouses providing the balance of the day's kindergarten education themselves, there was a vehement pushback. Spouses had given up career opportunities to spend quality time with their young children.

The superintendent wanted to know what the best research said about this situation and asked if I could assemble a group of faculty members to deliberate with him and his administrators in arriving at a decision regarding day-long kindergarten. This I did, reaching across the Stanford campus for our best relevant scholars. When we came together, the superintendent laid out the issue. He also described the research his staff had done so far, which included reviewing research literature but even more importantly, looking at the handful of districts around the country that had confronted a similar issue. What became clear is that decisions are situated and moral—the context surrounding the Palo Alto schools and what is best for children and families—and what others had done in similar situations. While the research his staff had reviewed and that we brought to the table were useful in a normative way, he and his administrators had to see how it played out in practice... on the ground. The discussions that ensued were enlightening on the academic and practice side. Ultimately the district went with full-day kindergarten. And the series of meeting expanded to include about 10 other superintendents working on common problems with our faculty.

This (and other) experiences led me to believe that:

Decisions about educational practice are complex. They involve multiple values and goals; are influenced by interacting social, historical, economic, and political forces; and always require trade-offs. In the best of all worlds, they are rational or at least reasonable. Rigorous empirical evidence should be brought to bear when available (even with the inherent limitations of inference), but it is not sufficient. Education-practice decisions are local, practical, moral, and made even though there is rarely enough evidence to clearly dictate a specific choice; they are always made with uncertainty. (Shavelson, Fu, Kurpius, & Wiley, 2014)

This said, subjectivity in educational decision-making (as it is most often constructed out of preference, inclination, and/or bias) should be avoided; rigorous empirical inquiry is one of the surest ways we have of diminishing subjectivity in decisions and action.

Closing Reflections

I believe that education is a practical enterprise, perhaps the most important enterprise any nation engages in. In all countries, but especially in the United States, education is based on a set of cultural beliefs and values. Being a diverse nation, our cultures and beliefs vary tremendously.

Our founding fathers were wise to make education a state and local enterprise with culture and subject-matter “wars” fought out in context, else the nation come crumbling down. Of course the federal government is involved and part of the wars surrounds how much involvement is appropriate and constitutionally legitimate.

But the “wars” are about beliefs and values. They are typically subjective. (One role of science is to evaluate and provide a basis for moving to objectively reasonable beliefs.) Moreover instead of laying bear just what those beliefs and values are and debating them as a democracy locally or nationally, we instead conduct the “wars” in the guise of research rigor—does inquiry teaching or direct instruction have the greatest impact on students’ academic achievement and did the research rule out all possible selection and other bias? That is we dodge the real beliefs and values wars under the guise of research rigor and invoke scientific rigor to what is ultimately a practical, value, and belief-ridden enterprise.

To be sure science and rigorous empirical education research can inform us of “what works” in a particular context at a particular time with a particular set of actors. More generalizable claims are suspect. But science cannot answer moral and normative questions. Scientific research can contribute evidence that challenges the tendency to rely on familiar experience, popular wisdom, and intuition in decision-making. But it is just part of the practical equation. Practitioners and policymakers have many other considerations to put into the equation including time, place, and contextual and moral dimensions of both the goals they set and the means by which they pursue them.

In the end, research and scholarship in education needs to draw on fields other than science to address questions of policy, practice, morals, and beliefs. Eisenhart’s listing (see above) is a good start. I would, as Cronbach did, include the humanities and arts as well. Stories and other depictions of life (including education!) are important. We learn a lot from them. These stories and depictions when put in practice can be examined for their moral, belief, and scientific justification. Scientific research in education can contribute to the empirical evaluation of the viability of alternatives. It along with interpretative science, philosophy, history, ethics, and critical research can influence and even change the “mindframes” of decision-makers and the public—evaluating what is proposed and creating an image of what might be possible (Shavelson, 1988). Of course the voices will not sing in harmony. But in diversity is evolution (as Darwin told us!). Challenges and counter challenges clarify values, beliefs, and evidence and such debate is healthy and essential in a democracy.

Perhaps the greatest limitation of SRE, then, was not in what it said but in what it emphasized and what it said but did not emphasize. More likely its greatest limitation may have been in what it did not say—

Eisenhart and Towne's second question: Is scientifically based research the only or best approach? However, the NRC was not the place to take up the question. Perhaps the National Academy of Education is the appropriate place.

References

- Berliner, D. C., & Glass, G. V. (in press). Trust, but verify. *Instructional Leadership*.
- Eisenhart, M. (2005). Science plus: A response to the responses to *Scientific Research in Education*, 107(1), 52–58.
- Eisenhart, M., & Towne, L. (2003, October). Contestation and change in national policy on “scientifically based” education research. *Education Researcher*, 32(7), 31–38.
- Feuer, M. J., Towne, L., & Shavelson, R. J. (2002, November). Scientific culture and education research. *Educational Researcher*, 31(8), 4–14.
- National Research Council. (2002). *Scientific research in education*. Committee on Scientific Principles for Education Research. R. J. Shavelson & L. Towne (Eds.), Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Phillips, D. C. & Burbulus, N. C. (2000). *Postpositivism and educational research*. Lanham, MD: Rowman & Littlefield Publishers.
- Shavelson, R. J. (1988). Contributions of educational research to policy and practice: Constructing, challenging, changing cognition. *Educational Researcher*, 17, 4–11, 22.
- Shavelson, R. J., & Webb, N. M. (2009). Generalizability theory and its contribution to the discussion of the generalizability of research findings. In K. Ercikan & W.-M. Roth (Eds.), *Generalizing from educational research: Beyond qualitative and quantitative polarization* (pp. 13–32). New York: Routledge.
- Shavelson, R., Fu, A., Kurpius, A., & Wiley, E. (2014). Evidence-based practice in science education. In R. Gunstone (Ed.), *Encyclopedia of Science Education*. New York: Springer. Retrieved from <http://link.springer.com/referencework/10.1007/978-94-007-6165-0/page/7>.

Assessing and Improving Federal Education Research

Maris A. Vinovskis^{1,2}

Before the 1960s, the federal government provided only limited support for education research. But during the past 50 years, the federal government has expanded considerably the extent of its involvement in funding educational research. With the passage of Title IV of the Elementary and Secondary Education Act of 1965, for example, federal research funding to the Office of Education dramatically increased from \$3 million in fiscal year (FY) 1960 to nearly \$100 million in FY1967 (of which approximately \$40 million supported the regional education laboratories and the research and development [R&D] centers).

Over the years, federal education research has not been a high priority for most of the public and many policymakers. Some people assume that they already know what constitutes a good education and have definite ideas about how it should be improved. Rather than funding more education research, they are more interested in disseminating the existing good ideas and implementing them into the classrooms. Others, who may acknowledge the need for more education research, are sometimes skeptical of what we have actually learned from the monies already spent on federal education research. As a result, although federal research funding

¹ Maris A. Vinovskis is the Bentley Professor of History, Research Professor of the Institute for Social Research, and Faculty Member of the Gerald R. Ford School of Public Policy at University of Michigan. He was elected to the National Academy of Education in 1996.

² Rather than providing a detailed documentation of the numerous quotes and sources used in this short essay, the reader is referred to the two references listed at the end of this document.

for education has increased overall, many knowledgeable scholars and educators believe that it is still woefully inadequate for the task of providing a high-quality education for all of our children today.

Much has already been researched and written about the accomplishments as well as the shortcomings of federally funded education research. Indeed, members of the National Academy of Education (NAEd), both individually and collectively, have played a vital role in this area.

There is one issue that may warrant more attention: assessing the overall quality of the federally funded education research in the past as well as today. Individual scholars have commented in passing on the quality of past education research and many of them have suggested considerable room for improvement. What have been generally lacking are more systematic efforts to assess the quality of the federal involvement as well as how to help alleviate any shortcomings. Since the late 1980s onward, however, there have been scattered efforts to address this issue as illustrated by some of the experiences of the Office of Education Research Improvement (OERI) and its successor, the Institute of Education Sciences (IES). A brief review of some of them may suggest future avenues for encouraging federal research improvements and stimulate more funding opportunities.

Concerns about the quality of statistics produced by the National Center for Education Statistics (NCES) led to the creation of a National Academy of Sciences (NAS) panel to investigate the NCES in the mid-1980s. The NAS panel unanimously stated that unless immediate steps were taken to improve the poor quality of data, perhaps NCES should be abolished and alternative sources of education statistics found. Fortunately, NCES took the report seriously and made substantial improvements in its data collection and analysis.

In the early 1990s, OERI Assistant Secretary Christopher Cross commissioned the NAS to assess the role of the federal government in education research and development. OERI wanted the NAS panel to assess the quality of the work produced by the labs and R&D centers, but this was not done (although the panel noted the problems that OERI had with its quality assurance system). The panel made useful suggestions for reorganizing OERI and improving its work. The 1992 report added that the "innovative methods, programs, and processes developed by the centers should be subject to a quality assurance process before wide-scale distribution."

Reinforced by the NAS panel recommendations as well as other questions about the quality of educational research and development, OERI Assistant Secretary of Education Diane Ravitch recruited me to be the OERI research advisor in 1992. One of my major responsibilities was to review the quality of research and development produced by the labs and

R&D centers in recent years. I reviewed the quality of R&D produced by 5 of the current 20 R&D centers as well as 5 of the 10 regional education labs. Preliminary results were shared with each of the centers and labs, as well as other interested parties in 1992. The final report was released in 1993.

Although praising many aspects of the lab and center contributions, the final report presented a mixed picture of the overall quality of center work:

On the one hand, some of the products produced within centers have produced excellent social science research.... On the other hand, some of the research products are so conceptually and methodologically weak that they either should never have been funded or should have been promptly improved after being funded.

In terms of the regional education labs, I noted that "today, for all practical purposes, many of the labs are primarily regional institutions offering research-based technical assistance." I also had serious questions about the quality of the R&D projects produced at the five regional labs (with the notable exception of the Far West Laboratory).

During the late 1980s and early 1990s, Congress paid relatively little attention to the quality of federally funded education research. Increased scholarly attention to the issue of education research quality as well as 2 days of OERI reauthorization hearings before the House Subcommittee on Select Education in March 1992 raised congressional concerns about the agency's political objectivity and its lack of adequate quality assessment systems, Congress reauthorized the agency in 1994 and mandated that OERI improve and implement three standards:

1. Standards for the evaluation of applications for grants, contracts, and cooperative agreements;
2. Standards for reviewing and designating exemplary promising programs; and
3. Standards for evaluating the activities and products of all recipients of OERI financial assistance.

OERI was able to develop and implement the first two standards reasonably well. Given the diversity of the agency's products, the third standard was much more complex. Although OERI worked hard to comply with this mandate, it was unable to develop and implement an adequate in-house quality system for its products (e.g., the third-year reviews of the R&D centers and the regional education labs did not employ the congressionally mandated rigorous standards). However, the 1994 OERI reauthorization reinforced the importance of having such quality stan-

dards and alerted policymakers to the need to focus more closely on them in the future.

The issue of education research quality standards received more congressional attention when the Comprehensive School Reform bill passed in 1997, which emphasized the use of scientifically based interventions (although none of the 17 acceptable models cited in the legislation provided convincing scientific evidence of being able to raise low-performing student outcomes). Similarly Congress passed the GOP's Reading Excellence bill in late 1998. The GOP bill included a definition of scientifically based reading-research (which some opponents attacked as too narrow and inappropriate).

In anticipation of reauthorizing OERI in 2000, Chairman Michael Castle (R-DE) of the House Subcommittee on Early Childhood, Youth, and Families proposed the Scientifically-Based Education, Statistics, Evaluation, and Information Act of 2000, which would set up an independent agency. The proposed legislation also included a definition of high-quality research that was contested by some in the scholarly community. Disagreements over having an autonomous independent agency led to a compromise, but otherwise the House subcommittee adopted Castle's bill unanimously (although Senators postponed action on it until the following year).

As some scholars worried about the new congressional definitions of scientifically rigorous education research, OERI's National Educational Research Policy and Priorities Board commissioned the National Research Council to review how education research should be defined. The Committee on Scientific Principles for Education Research was created but decided

not [to] attempt to evaluate the quality of bodies of existing research of existing researchers in the field because that would have constituted a monumental challenge and we judged it beyond the scope of our charge. Instead, we adopted a forward-looking approach that draws on lessons from history and identifies the roles of various stakeholders (e.g. researchers, policymakers, practitioners) in fulfilling a vision for the future of education research.

The Committee on Scientific Principles for Education Research report provided six scientific principles that should be followed in providing high-quality research. The report stressed the responsibility of the community of education researchers to improve the quality of education research, but opposed the recent congressional efforts to define and mandate specific scientific standards through legislation. The report was widely discussed by policymakers and educators and generally well-received, although questions were raised about its particular definitions

of scientific rigor as well as how the standards were to be implemented in practice.

The No Child Left Behind Act was passed in 2001 by the 107th Congress. Although Republicans and Democrats disagreed on some provisions of the legislation, they both endorsed the necessity for using scientifically based or research-based programs and practices. Although there was little discussion of education research by Congress, those terms were used more than a 100 times in the final legislation, including the development of a rigorous definition of scientifically based research. Some outside commentators questioned the availability of such research today, while others thought too much emphasis was placed on quantitative methods and randomized field experiments. Educators were also concerned about how these research standards would be implemented and administered. Unfortunately, the U.S. Department of Education did not provide much clarification for local and state officials.

With the passage of No Child Left Behind in 2001, Congress turned to the reauthorization of OERI. Castle had introduced the Education Sciences Reform Act of 2002 and held hearings on July 17, 2001, and February 28, 2002. At those hearings, there were differences of opinion about the quality of education research.

The National Research Council's panel released its 2002 report, *Scientific Research in Education*, and acknowledged that the quality of education research was unequal, but so was the research in other areas. The report stated that

the conventional wisdom about the weakness of scientific educational research relative to other sciences is exaggerated, and the criticisms would be equally worthy of serious investigations if leveled at other branches of the social and physical sciences or at other applied fields like medicine or agriculture.

Other witnesses were more critical of the quality and usefulness of education research, Frank Newman, professor of public policy and sociology at Brown University and president of the Education Commission of the States, observed that "the problem with research, at least public scholarship, is not a deficiency of quantity, but of quality. The problem is that research in this country is grossly inadequate to the task." Jim Horne, Secretary of the Florida Board of Education, added that

there is a broad consensus today at the state and local levels that much of the research that has been funded and disseminated by the Federal government has not, to date, met the same very rigorous and stringent criteria that is now defined clearly in the No Child Left Behind law.

With relatively little disagreement between Democrats and Republicans, the Education Sciences Reform Act of 2002 was signed into law and the George W. Bush Administration created the IES. The new agency, under Assistant Secretary Grover “Russ” Whitehurst made improvements such as creating the What Works Clearinghouse (which employed more rigorous standards) and financing randomized education experimental projects.

While I have followed fairly closely efforts to assess the quality of educational research and development in the past as well as in the early 2000s, I have not kept as close a look at this issue in more recent years. Certainly the National Academy of Education has shown interest in this area as evidenced by its panels on these topics at the annual meetings (on which I have occasionally participated).

I wonder, however, that given continued questions about the quality and usefulness of education research as well as the growing concerns of policymakers for improvements in this area, perhaps this might be the time for the NAEd and others to look more systematically and carefully at these questions. With the reauthorization of No Child Left Behind and IES coming as well as the NAEd’s continued commitment to high-quality education research, maybe it would be worthwhile to assess what has been tried and accomplished during both the Bush and Obama administrations in improving the scientific quality of education R&D during the past 15 years.

References

- Vinovskis, M. A. (2001). *Revitalizing federal education research and development: Improving the R&D centers, regional educational laboratories, and the “new” OERI*. Ann Arbor, MI: University of Michigan Press.
- Vinovskis, M. A. (2008). A history of efforts to improve the quality of federal education research: From Gardner’s Task Force to the Institute of Education Sciences. In P. B. Walters, A. Lareau, & S. H. Ranis (Eds.), *Education research on trial: Policy reform and the call for scientific rigor* (pp. 51–79). New York: Routledge.

Improvement of Early Childhood and K–12 Education

Empowering Our Best Teachers: Essential for Producing More Effective Systems of Education in the United States

*Bruce Alberts*¹

My Long Involvement with K–12 Education Issues

The title of this essay presents a major conclusion that I have reached after three decades of working closely with U.S. public school districts. My interest in improving school system management began in the early 1980s, when my wife, Betty Alberts, became the president of the San Francisco Parent Teacher Association (PTA). Her new position required that I listen to the meetings of San Francisco's elected school board, where she would often speak. The discussions there shocked me because very little attention was paid to fundamental education issues. Nor was there any obvious way for the district's best teachers and principals to provide the board with the kind of information that it clearly needed to govern wisely.

Since then, I have been involved in efforts to improve the science education experiences for students in grades K–12, including serving as the principal investigator during the early 1990s for a major National Science Foundation (NSF) grant for elementary school science in San Francisco, called City Science. More recently, a close involvement with the National Academies' Teacher Advisory Council and the California Teacher Advisory Council has provided me with many meaningful interactions with

¹ Bruce Alberts is the Chancellor's Leadership Chair in Biochemistry and Biophysics for Science and Education at the University of California, San Francisco. He was elected to the National Academy of Education in 2003.

some of our nation's most outstanding science and mathematics teachers.² In addition, since 2005, I have served as the board chair for the Strategic Education Research Partnership (SERP), a nongovernmental organization (NGO) that works closely with a set of school districts that serve as field sites for its research and development efforts (National Research Council, 2003; see also <http://www.serp.institute.org>).

The Dysfunctional, Top-Down Management of U.S. School Systems

Long ago, U.S. business learned the benefits of constantly soliciting advice from workers on the shop floor by studying the startling success of the Japanese automobile industry. However, the vast majority of U.S. school districts, failing to adjust to this fact, remained top-down, hierarchical operations. To make matters worse, the federal government's No Child Left Behind Act of 2001 introduced a heavy-handed emphasis on test-based accountability, with sanctions for failing schools. These top-down demands on U.S. K–12 education systems have exacerbated the “command and control” tendencies in school districts. The terrible consequences are articulately expressed by an East Coast leader in science education and outreach:

We are currently working in several of the schools, and here's my assessment of what I have seen. Compliance has replaced a focus on learning. What you hear from the very well-intentioned people who work in the neighborhood schools is a version of: “you can't believe all the things we have to do.” They see themselves less as individuals who are trying to help kids learn, and more as victims of a system that is ordering them about and oppressing them. Strong words, I know, but I find it stunning that we have turned too many of our schools (particularly poor, urban ones) into such spectacular messes. From the principals to the teachers, no one feels entitled to exercise their authority to think how they can use what they are being asked to do to effect meaningful change.³

What this means for our teachers is reflected in a Finnish high school teacher's summary after his recent visit to U.S. schools:

Well, a surprise for me was in States that I have heard many, many stories about how bad the teachers are in the U.S.... But those four months I was there and I traveling through many, many classrooms in that time, and I didn't see any bad teachers. But I saw teachers that work way more than I do.... Teachers in the U.S. have to work too much, I wouldn't say

² See <http://sites.nationalacademies.org/dbasse/tac/index.htm> and <http://ccst.us/ccstinfor/caltac.php>.

³ Personal communication, April 27, 2015, Margaret Honey, President and CEO, New York Hall of Science.

nonsense, but too much on what doesn't help teaching or doesn't help learning—Lots of reports, lots of meetings with no goal, and maybe meetings just for meetings, and also reports on students. You know, “they have done this and they have done this.” They work hard, but not with the students. They work hard with the system. And that was the biggest difference in our educational systems. (English, 2014, Chapter 18)

My daughter Beth Alberts, a high school science teacher, reports that nearly everyone in her school district “works hard with the system”—not only the teachers, but also the principals and the central office staff. There are too many regulations and forms to fill out. Everyone blames someone else for the “messes”—the principals blame the district, the district blames the state, and the state blames the federal government.

From the Top Down: Sins of Commission and Sins of Omission

The bureaucratic burdens in U.S. school districts can be classified into two categories: sins of *commission* and sins of *omission*.

The sins of commission consist of demands placed on schools and teachers that interfere with student learning, either through destructive requirements (e.g., a month of test preparation each spring) or through destructive rules that prevent teachers from teaching well, as noted in the example below.

As a cell biologist, I claim that a living cell is the most amazing thing that we know about in the universe: a tiny sophisticated chemical system that can replicate itself indefinitely. All students should experience the wonder of the living cell in science class. For decades, a standard experiment for 12-year-olds has had them rub the inside of their cheeks with a Q-tip; they then transfer the material picked up by the cotton onto a glass slide, allowing them to examine some of their own cells in a microscope. In California, this harmless experiment cannot be done without obtaining signed parental permission for each student involved. To further discourage this bit of active science, each student's slide must be discarded in a toxic waste container, which the school district is required to dispose of specially. No one seems able to explain the rationale for either of these two requirements. However, these restrictions, and many others like them, help to explain why middle school students generally find cells boring. To protect them from their own cheek cells, they instead memorize what a cell looks like from drawings in their textbook. This is an example of what I call a sin of commission.

My example of a sin of omission likewise comes from science education, where materials and supplies are required for the inquiry-based, active science learning that has long been called for in our nation's schools (American Association for the Advancement of Science, 1986; National

Research Council, 1996, 2012). In 2007, the San Francisco Unified School District (SFUSD) purchased the Full Option Science System (FOSS) units for all of its elementary school teachers. However, it neither provided adequate professional development for the elementary teachers in its approximately 70 elementary schools, nor restocked the consumable supplies in the FOSS kits after they were used. As a result, much of the investment made in these high-quality science units was wasted. I attribute this failure to the fact that the top district leaders were unaware of the problems, being far too insulated from what actually happens in the schools.

A Systemic Undervaluing of Teachers and Their Expertise

Ronald Thorpe, in his important article “Sustaining the Teaching Profession,” wrote the following after meeting with a group of our nation’s best teachers, who had recently completed a year in Washington, DC, as Einstein Distinguished Educator Fellows.⁴

They were not looking forward to returning to their schools and classrooms. Why? Because they had just experienced—perhaps for the first time in their professional lives—what it is like to be treated as a real adult with real knowledge, skills, and opinions.... They would never receive such respect back in their schools, where they might even encounter resentment from colleagues and administrators. What a loss! (Thorpe, 2014, p. 15)

The bottom line is that if our schools cannot reabsorb the handful of Einstein Fellows and give them more responsibility for improving teaching and learning, there is no hope for our profession and our schools.... Everyone involved with schools and districts must find ways to use the talent they have among their teachers to the greatest advantage. Holding them in lock-step positions forces the best people out of the profession and undoubtedly convinces many people not even to explore the possibility of becoming a teacher. (Thorpe, 2014, p. 15)

How can we change the culture of schooling so that it becomes routine and expected that outstanding teachers will provide effective, regular input to help steer their school district’s (and their state’s) policies and practices? School districts cannot succeed as strictly hierarchical organizations with a “command and control” mode of operation. Unless the United States can make serious inroads on this problem, we will never have public school systems that make the best decisions for their students. Nor will we be able to attract and retain a talented teacher corps. (Thorpe, 2014)

⁴ See <http://www.trianglecoalition.org/einstein-fellows>.

Current Attempts to Make Better Use of Teacher Expertise

In recent years there has been an increasing, widespread recognition of an urgent need to do much more to empower our best teachers. A 2015 book, *The Cage Busting Teacher*, contains a list of more than 30 organizations with such aims (Hess, 2015). Most seem focused on creating a cohort of lead teachers who use their skills to improve their own schools, through mentoring sets of their teacher colleagues and/or through various forms of distributed school governance (see Berry & Byrd, 2013; Hess, 2015; Valdez & Broin, 2015, for a range of such efforts; see also <http://www.teachingquality.org> and <http://www.teacherpowered.org/about>).

Especially notable is a push for “teacher-led schools” with distributed leadership—schools in which the teachers select their leaders, select their colleagues, and set schedules—while also determining staffing patterns, the learning program, and school-level policies. This form of management makes great sense to me, resembling the way that departments are managed in universities. For example, I find it amazing that school principals so often act alone in hiring new teachers for a school. Sadly, most current school leaders appear to lack the skill set and vision needed to build the collaborative, team culture required for real school improvement (Talbert, 2010).

The current attempts to improve schools by treating teaching as a true profession will be important for improving the quality of education that students receive in our schools. However, I find them insufficient. We also need a much stronger focus on harnessing the wisdom of lead teachers to continuously improve our local, state, and national education systems.

The Urgent Need for a Change in School System Culture

My personal sense of frustration was forcefully expressed by Alfred North Whitehead:

The art of education is never easy. To surmount its difficulties, especially those of elementary education, is a task worthy of the highest genius.... [But] when one considers ... the importance of this question of the education of a nation's young, the broken lives, the defeated hopes, the national failures, which result from the frivolous inertia with which it is treated, it is difficult to restrain within oneself a savage rage. In the conditions of modern life the rule is absolute, [a country] that does not value trained intelligence is doomed. (Whitehead, 1929, p. 14)

The urgent need for action in harnessing the “trained intelligence” of our many outstanding teachers leads me to raise two important questions. Might a coalition of organizations across the United States launch a

movement to create a new expectation with respect to school system governance in the United States? If so, then what might its initial strategy be?

To have any chance of altering the deeply embedded, hierarchical culture of today's school districts, any new expectation must not require substantial restructuring of the bureaucracy. Furthermore, to be effective, it should not completely remove the lead teachers involved from their classrooms, converting them into "bureaucrats" who are viewed as being out of touch with reality by their colleagues.

A Possible Strategy for Improving Education Through Systemic Teacher Leadership

Given the above constraints, what might be a successful strategy for empowering teachers in a way that openly and explicitly uses their wisdom of practice to improve the effectiveness of school systems? This essay ends with some suggested ingredients of such a plan, focused at the school district level and presented as a series of possible steps.

1. The teacher empowerment process would begin with an announcement by the superintendent and/or the district's school board that a small group of outstanding teachers will be specially selected to serve in an advisory role. This "Teacher Advisory Group" would be chartered to provide the superintendent (and board) with honest feedback from the district's classrooms. The group's findings and advice on how to adjust school district services and policies to improve the education of students would be disseminated as public information.
2. The critical next step would be careful selection of a small set of lead teachers for each school district (perhaps 10 teachers, depending on district size). These experienced individuals would continue their teaching for 50 percent time, while being paid to perform meaningful, non-bureaucratic leadership roles for the remaining half time. To ensure a strictly merit-based selection and strong credibility, these teachers should have received some type of outside recognition, such as National Board Certification. Ideally, they would be selected by a panel that includes representatives of some of the district's local partner organizations (e.g., colleges, parent teacher associations, business groups, and/or education NGOs). Each lead teacher would be appointed for a fixed time period (e.g., 3 years), with overlapping terms that ensure rotation.
3. The range of issues to be considered by this new Teacher Advisory Group would need to be specified in writing to make sure

that it focuses on the appropriate issues. It is critical that the group not be distracted by issues covered elsewhere, such as teacher compensation (addressed by the teachers union), the closing or restructuring of specific schools (addressed by the school board), and so on. To help catalyze a national movement, a set of model charters for such a group should be produced and widely promulgated by interested experts—including teachers unions and NGOs involved in education.

4. Although personnel in the district central office will be important sources of advice and technical support, the Teacher Advisory Group must not be viewed by colleagues as just another part of central office bureaucracy. To this end, for the entire range of issues included within their remit, the group should be empowered to determine its own mode of operation, as well as the priority issues that its members will address each year. The group should be empowered to elect its own officers and to divide tasks appropriately (e.g., chair, vice chair, secretary, communications, website maintenance).
5. To ensure its relevance and credibility, this Teacher Advisory Group will need to reach out energetically to other teachers in the school district, periodically soliciting their input and feedback, while constantly keeping them informed with regard to its activities. However, it is important that the group use its own judgment to lead—avoiding a mere reporting on average teacher opinions, for example.
6. Because much of what is being suggested represents new territory in U.S. education, a vigorous, high-quality research effort should be launched to study the successes and failures of this new national effort to provide useful guidelines for improving such teacher empowerment processes in the future. Hopefully members of the National Academy of Education would play an important role in such efforts.

In the space allotted for these essays, I have not found room to discuss mechanisms for incorporating more teacher wisdom into state and federal policymaking. However, a much louder voice is needed for our best teachers at these higher levels as well. For science and math education, a start has been made at the national level with the annual appointment of Einstein Distinguished Educator Fellows and the establishment of the National Academies' Teacher Advisory Council. Likewise, the California Teachers Advisory Council represents a start at the state level (see Footnotes 2 and 4). However, in the future even more effective ways to engage

with policymakers at both the state and national levels will need to be developed.

References

- American Association for the Advancement of Science. (1986). *Science for all Americans*. Washington, DC: Author.
- Berry, B., & Byrd, A. (2013). *Teacherpreneurs: Innovative teachers who lead but don't leave*. San Francisco, CA: Jossey-Bass.
- English, J. (2014). Finns share their insights about American education. In *The "Finnish way" to optimize student learning*. Retrieved from <http://www.TheFinnishWay.org>.
- Hess, F. M. (2015). *The cage-busting teacher*. Cambridge, MA: Harvard Education Press.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2003). *Strategic education research partnership*. Washington, DC: The National Academies Press.
- National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Washington, DC: The National Academies Press.
- Talbert, J. A. (2010). Professional learning communities at the crossroads: How systems hinder or engender change. In A. Hargreaves, A. Lieberman, M. Fullan, & D. Hopkins (Eds.), *Second international handbook of educational change*. Heidelberg, Germany: Springer-Verlag.
- Thorpe, R. (2014). Sustaining the teaching profession. *New England Journal of Public Policy*, 26(1), Article 5. Retrieved from <http://scholarworks.umb.edu/nejpp/vol26/iss1/5>.
- Valdez, M., & Broin, A. (2015). *Untapped: Transforming teacher leadership to help students succeed*. New York: New Leaders. Retrieved from http://www.newleaders.org/wp-content/uploads/NewLeaders_Untapped.pdf.
- Whitehead, A. N. (1929). *The aims of education and other essays*. New York: The Free Press.

Vannevar Sets the Stage

*Richard C. Atkinson*¹

This essay is a personal account of some events in my life associated with the establishment and evolution of the National Academy of Education (NAEd). I begin with some comments on U.S. science policy and conclude with a brief account of my own research on the educational process.

The Bush Report

Near the end of World War II, President Roosevelt—recognizing science’s remarkable contributions to the war effort—asked his science advisor, Vannevar Bush, to define a plan for science in the post-war era. That request led to Bush’s report *Science: The Endless Frontier* (Bush, 1945). What was the nature of that report? No summary could do justice to Bush’s masterful analysis, but essentially he made three principal arguments about the future of the U.S. scientific enterprise. First, he argued that most aspects of research and development (R&D) are the responsibility of the private sector. However, he also recognized that market mechanisms discourage the private sector from investing adequate funds in basic research. This recognition led Bush to his second argument: ensuring support for basic research in the post-war period should be the responsibility of the federal government, because the enormous benefits to society at large justify the investment. He believed that basic research should be

¹ Richard C. Atkinson is President Emeritus of the University of California. He was elected to the National Academy of Education in 1974.

conducted in the universities of the nation, rather than in government laboratories. As the institutions responsible for the nation's basic research, universities had pride of place in Bush's vision of the research enterprise. Third, he argued that decisions about which university research projects should receive government funding should be made via a peer-review process.

The Bush report remains to this day the single most important document on U.S. science policy ever written and a landmark for federal legislation. Before World War II, the federal government provided virtually no funds for research at universities; the very concept was viewed as radical. In the post-war period, the government committed itself to becoming the principal sponsor of scientific research to be conducted primarily at universities. It was an extraordinary reversal of direction.

The Bush report led to the establishment of the National Science Foundation (NSF) and the reorganization of the National Institutes of Health (NIH) and other federal agencies to support extramural research. Initially, the flow of funds for research moved at a slow pace and was primarily focused on the physical and biological sciences and engineering. However, when the Soviets successfully launched the satellite Sputnik in 1957, America began to question its leadership in science. It responded with a massive increase in funding for science to include the behavioral and social sciences. In addition, there was a sense that the United States had fallen behind the Soviet Union in science education, particularly in grades K–12. The response was a series of large-scale curriculum projects principally funded by NSF. These projects enlisted some of America's most famous scientists who worked collaboratively with educators to develop curriculum. The projects proved to be successful, but it soon became evident that the body of research to guide the effort was insufficient. The curriculum projects and related federal ventures in education led to a major expansion of the educational research enterprise that involved more funding for research and fellowships to attract individuals of outstanding ability. The field of educational research blossomed during this period and involved a mix of scientists from various disciplines, including those whose first identity was as an educational researcher.

Formation of the NAEd

To ensure the future of educational research, it was time to establish an academy of individuals elected on a national/international basis for outstanding scholarly and research contributions relevant to education. In the early 1960s, Francis Keppel, U.S. commissioner of education, began an exchange with Ralph Tyler, director of the Center for Advanced Study in the Behavioral Sciences, about a way to evaluate the state of U.S. edu-

cation. That exchange was followed by a series of committee meetings under the sponsorship of the Carnegie Corporation of New York that laid the foundation for the National Assessment of Educational Progress and increased federal funding for educational research. During this period it became evident that an organization such as the National Academy of Sciences (NAS) could play a facilitating role. John Gardner, president of the Carnegie Corporation, then wrote a letter inviting a leading group of individuals involved in education to form the NAEd, and Carnegie provided start-up funding to make it a reality.

In the 1970s, when I was NSF director, I brokered a number of meetings between the NAEd leadership and Philip Handler, NAS president, with the goal of including the NAEd under the umbrella of the NAS/National Research Council. These talks were cordial, and Handler always expressed high regard for the NAEd members. However, there were too many obstacles to overcome, not the least of which was the precarious state of the NAEd finances and the absence of any endowment. The NAEd's current president, Michael Feuer, has engaged in similar discussions with the NAS leadership in recent years. There has always been a desire to cooperate, but not as yet to join forces.

Evolution of Computer-Assisted Instruction

My involvement with educational research began in the late 1950s as a newly appointed member of the faculty at Stanford University; it was an unusual joint appointment involving the Department of Psychology, School of Education, School of Engineering, Statistics and Applied Mathematics Laboratories, and Institute for Mathematical Studies in the Social Studies. In those days my main research interest involved the formulation and testing of mathematical models for learning and conditioning in both humans and animals. Patrick Suppes, a charter member of the NAEd, was my colleague at Stanford. He was a professor of philosophy with a special interest in mathematical logic and the philosophy of science. He was a leading authority on the role of formal models in the development of scientific theories. Suppes was one of those individuals from another discipline who joined the NSF effort to develop the new mathematics curriculum; his own work was on mathematics in grades K–3. Given my interests in models of memory and learning, his work on the role of models in science, and his newly formed interest in how young children acquire mathematical skills, we collaborated on a number of studies.

In 1962, Suppes and I received a grant from the Carnegie Corporation to support the use of a computer to conduct psychological experiments. Of special interest was the idea of teaching reading and mathematics to young children under computer controls with the capability of individu-

alizing the instruction. We purchased a PDP-1 computer manufactured by Digital Equipment Corporation; it was one of the first transistorized computers. We quickly had six terminals running on a time-sharing system and were busing kindergarten and first-grade students to our laboratory at Stanford. Encouraged by our initial success, we applied and received a \$1 million grant from the U.S. Office of Education (this was before the U.S. Department of Education existed). In those days \$1 million grants were rare; even the physics community took note.

Our plan was to develop a computer assisted instruction (CAI) system to teach reading and mathematics to culturally disadvantaged, K–3 children. Our group at Stanford undertook the design and implementation of what became known as the IBM 1500 Instructional System. The 1500 system was installed at a school in East Palo Alto and went into operation in the fall of 1967. The 1500 system was housed in two trailers on the school parking lot. One trailer housed the computer system; the other trailer, 16 student terminals. Each student terminal had a cathode-ray tube (CRT), a typewriter keyboard, a light pen to touch a point on the face of the CRT, a projector with a capacity of 1,000 color images, a set of earphones with a microphone, and pre-recorded audio messages that could be “randomly” accessed (this was before digital audio was commercially available). Suppes had responsibility for developing the mathematics curriculum, and I had responsibility for the reading curriculum.

By the end of 2 years, approximately 400 students had received a major part of their daily instruction in reading and mathematics under computer control. As the first installation of its kind, the system received considerable national attention; more than 3,000 visitors per year had observed students at work on the system. More importantly, significant gains in student achievement had been demonstrated. A description of our work with the 1500 system is available in an article titled “Computerized Instruction and the Learning Process” (Atkinson, 1968).

The 1500 system permitted us to individualize the learning process, but not to the extent we wanted. The IBM 650 drove the system, which was the first computer to be widely adopted by American universities; today’s iPhone has 10,000 times the computing power of the IBM 650. Furthermore, the system’s cost was prohibitive, and locating the computer at the school site had major disadvantages. Fortunately, while working with the 1500 system, we continued to expand the PDP-1 system housed at Stanford. The student terminals were simpler: a low-cost display device, a typewriter keyboard, and a headset supported by digital audio that was truly random access. We soon had about 40 terminals in several Stanford buildings connected to the computer by phone lines. It was not a big step to connect to schools at remote sites. We restructured the reading and math programs for the Stanford system, and by 1967 about 3,000 students

were receiving daily instruction in seven nearby elementary schools and in locations as distant as McComb, Mississippi; Morehead, Kentucky; and Washington, DC. The system and its effectiveness are described in an article titled "Teaching Children to Read Using a Computer" (Atkinson, 1974; Fletcher & Atkinson, 1972).

As the Stanford system was upgraded and enhanced, it was possible to experiment with a wider range of courses. Suppes developed a program in logic that he used to supplement his regular Stanford lecture course in introductory logic. My group developed a course in computer programming using the BASIC computer language, which was widely used by Stanford graduate and undergraduate students and at two local community colleges (Barr, Beard, & Atkinson, 1975). These courses were adaptive in two ways: (1) the sequence of instruction varied as a function of a student's performance history and (2) the CAI program could self-modify as more students completed the course and their data were used to update estimates of parameters that specified problem difficulty (Atkinson, 1976).

A principal goal of our CAI research was to experiment with different approaches to optimizing student performance. For some topics, we were able to formulate mathematical models of the learning process and then use methods of control theory to make moment-by-moment decisions about what should be learned next to optimize the student's performance. Several parts of the K-3 reading program and of the foreign language vocabulary programs provided elegant examples of this approach. In other cases, the "optimal" schemes were not optimal in a well-defined sense, but they were based on our intuitions about learning and relevant laboratory experiments. Elsewhere, I have used the term "theory of instruction" to describe the issues involved in using a theory of learning, formal or not, to develop an optimal program of instruction (Atkinson, 1972a, 1972b; Atkinson & Paulson, 1972; Chant & Atkinson, 1978; Groen & Atkinson, 1966).

During fall 1974, I was invited to be a visiting professor at Rockefeller University for the academic year 1975-1976. Part of my plan for the year was to write a book reviewing our research on CAI. The tentative title was *Theory of Instruction*. However, at the last minute my world changed. I was recruited to NSF, expecting to spend my sabbatical year in Washington, DC. I never returned to Stanford. My career as an active researcher ended at that time (Atkinson, 1999).

Since I left the field of educational research, the development of CAI has continued, and there are beautiful examples using psychological theory to individualize instruction. A variety of commercial entities, both large and small, have promoted the use of CAI in schools and universities and for training personnel in the military and corporate sectors. The

deployment of CAI has not been as rapid as I predicted in a 1969 article in the *Proceedings of the National Academy of Sciences of the United States of America*, but it has been substantial (Atkinson, 1969). Suppes was the most persistent and long-term contributor to the field. His efforts were truly remarkable, both in the development of new programs and in the detailed experimental evaluations of student performance. The Stanford University Online High School is an example of what he accomplished. This online, fully accredited, diploma-granting program for grades 7–12 serves students around the world. It has been in operation for over a decade with excellent results. Unfortunately, Suppes passed away in November 2014. He was a giant in the world of academia, and his death is a great loss to the field of educational research.

The world of CAI underwent a total transformation in 1994 with the advent of the Internet, which offers an instruction platform with a rich multi-sensory surround and a virtually unlimited computing capacity. Wireless communication has also contributed to this transformation; the flexibility of not being tied to the Internet by a cable makes a substantial difference in education. Since 1994 MOOCs (Massive Online Open Courses) and related efforts have been introduced. That work is interesting, but the key to success is individualizing instruction, which requires a theory of the learning process.

Conclusion

I conclude with a comment about education policy and the NAEd. From my personal and professional journeys, I have learned many lessons about the involvement of researchers in the formulation of policy. The NAEd plays a special role, as an organization focused on building and sustaining connections between scholarship and action, cultivating future researchers oriented toward the improvement of educational policy and practice, and providing a home for nonpartisan explorations of the basic and applied sciences of teaching and realities. Given the imperative of investing wisely in the development of human capital, it is more important than ever that our NAEd be a prominent and visible player.

References

- Atkinson, R. C. (1968). Computerized instruction and the learning process. *American Psychologist*, 23, 225–239.
- Atkinson, R. C. (1969). Computer-assisted learning in action. *Proceedings of the National Academy of Sciences of the United States of America*, 63, 588–594.
- Atkinson, R. C. (1972a). Ingredients for a theory of instruction. *American Psychologist*, 27, 921–931.

- Atkinson, R. C. (1972b). Optimizing the learning of a second-language vocabulary (1972). *Journal of Experimental Psychology*, 96, 124–129.
- Atkinson, R. C. (1974). Teaching children to read using a computer. *American Psychologist*, 29, 169–178.
- Atkinson, R. C. (1976). Adaptive instructional systems: Some attempts to optimize the learning process. In D. Klahr (Ed.), *Cognition and instruction*. Hillsdale, NJ: Erlbaum Associates.
- Atkinson, R. C. (1999). The golden fleece, science education, and U.S. science policy. *Proceedings of the American Philosophical Society*, 143(3), 407–417.
- Atkinson, R. C., & Paulson, J. A. (1972). An approach to the psychology of instruction. *Psychological Bulletin*, 78, 49–61.
- Barr, A., Beard, M., & Atkinson, R. C. (1975). A rationale and description of a CAI program to teach the BASIC programming language. *Instructional Science*, 4, 1–31.
- Bush, V. (1945). *Science the endless frontier*. Washington, DC: U.S. Government Printing Office.
- Chant, V. G., & Atkinson, R. C. (1978). Application of learning models and optimization theory to problems of instruction. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, NJ: Erlbaum Associates.
- Fletcher, J. D., & Atkinson, R. C. (1972). Evaluation of the Stanford CAI program in initial reading. *Journal of Educational Psychology*, 63, 597–602.
- Groen, G. J., & Atkinson, R. C. (1966). Models for optimizing the learning process. *Psychological Bulletin*, 66, 309–320.
- Pelfrey, P. (2012). *Entrepreneurial president: Richard Atkinson and the University of California*. Berkeley and Los Angeles, CA: University of California Press.

In Search of Broader Visions: An Appreciation and Future Hope for the National Academy of Education

*William Damon*¹

In recent years, public education policies in the United States have been marked by noble goals pursued by narrow and unimaginative means. Although brilliant work is being done in labs and schools all over the world, such work has been largely ignored by the dominant agencies that determine public policy in education. Prevailing policies properly aim to eliminate inequality in educational opportunities among students and to increase their skills in ways that will contribute to our economy and ensure their employability: these are noble goals. But the policies have been pursued through strategies that treat all students as if they have similar interests, aspirations, and personal qualities. What's more, the policies neglect elements of learning essential for students' intellectual and social development, and they are guided by measures that fail to capture many of the capacities that enable students to succeed in work and life.

Ironies and contradictions abound. Among our driving public policy priorities are student employability and national economic health (as measured by gross domestic product); yet, art and music have been marginalized in public education today, overlooking the economic fact that the entertainment and communication industries, which draw heavily on artistic talents, are among today's top sources of employment and productivity. Another policy priority is inequality among students; yet civ-

¹ William Damon is Professor of Education and Director of the Center on Adolescence at Stanford University. He was elected to the National Academy of Education in 1999.

ics education, which builds the capacity to achieve equal rights through political participation in the democratic process, has been neglected in favor of a single-minded focus on rudimentary skills.

This is not an untold story that needs breaking: the point has been clearly made through social commentary as well as solid research. In eloquent recent writings, National Academy of Education (NAEd) member Diane Ravitch (2010) has criticized the narrowness of present-day public school teaching and testing regimens, setting the terms of a vigorous national debate. In 2012, journalist Paul Tough published a best-selling book, *How Children Succeed* (Tough, 2012), emphasizing the importance of “non-cognitive skills” that public schools have been busy ignoring in our time. Studies by Angela Duckworth (2005) have reported that character strengths such as grit and self-control predict achievement more powerfully than do I.Q. and other cognitive tests. Research from a number of labs, including our own at the Stanford Center on Adolescence, has shown the importance of qualities such as diligence, curiosity, creativity, and purpose for both learning and accomplishment. Yet although such news has been available for the better part of a decade, it has not noticeably budged the big-time governmental and funding agencies that set and enforce our public education policies.

The part of this story that I shall discuss in this short essay is the diminishment of civics education in public schools; and, although this is an especially disturbing part of the story, there has been, in my own experience, a salutary NAEd dimension to it.

As with all the other capacities that our public schools have been forced to neglect by the prevailing public policies, the decline in students’ civic capacities is well known; and it has been rued. Sandra Day O’Connor has called this a “crisis in civics education” (Dillon, 2011). A National Assessment of Educational Progress (NAEP) report found only one in four high-school students scoring “proficient” in knowledge of citizenship. According to the NAEP report, “a smaller proportion of fourth and eighth graders demonstrated proficiency in civics than in any other subject the government has tested ... except history, American students’ worst subject” (National Center for Education Statistics, 2011). A Southern Poverty Law Center study (2011) found that, less than 50 years after its great successes, the civil rights movement is now rarely taught and little-known among students. An American Enterprise Institute study found that social studies teachers doubted that students understood core U.S. citizenship concepts such as the Bill of Rights and the separation of powers (Schmitt et al., 2010). A U.S. Department of Education study found that only 9 percent of high school students could cite reasons why it is important for citizens to participate in a democracy and that only 6

percent were able to cite reasons why a constitution benefits a country (National Center for Education Statistics, 2011).

Not only is the neglect of civic education in public schools ultimately self-destructive for any democratic society that must rely on educated citizens for its future, but also it radically departs from our own educational history. Thomas Jefferson wrote that cultivating civic virtue should be “the principal business of education.” Washington and Madison proposed national universities that would teach good citizenship to America’s younger generations. For much of the following two centuries, civics was a centerpiece of American schooling. In Tocqueville’s visit to America in the 1830s, he observed that the “general thrust” of American schooling “was directed towards political life,” unlike in the fading nations of Europe. When waves of new immigrants reached U.S. shores in the late 19th and early 20th centuries, the explicit mission of our public schools was to teach them how to be citizens in a democracy. A report from the Annenberg Institute at Penn noted “until the 1960’s, three courses in civics and government were common in American high schools: *Civics*, *Democracy*, and *U.S. Government*” (Gould, 2011). Since then, the report wrote, there has been a “decades-long decline” in civics instruction, fueled by increasing pressure on schools to raise student scores on tests of basic skills.

My own interest in youth civic development goes back to 1999, when the research center that I direct at Stanford hosted a conference on the topic. At that time, and for most of the period since, the operative concept driving research in this area was “engagement;” the problem most researchers worried about was an increase in “civic disengagement.” The problem was defined by data such as decreasing voting rates, in sharply linear fashion, among 18- to 24-year-olds. However, I was never convinced that the notion of disengagement captured the problem sufficiently: for one thing, diverse groups of youth seem engaged in civic activities that do not show up in traditional measures; and for another thing, engagement is too vague a concept to define the actual cognitive and behavioral processes involved in civic participation. For a more effective approach to civic education, I believed, more powerful conceptual handles were needed.

This is when, in my experience, the NAEd played a constructive role. The first step was a conversation I had with James Banks, one of the great founders of the civic education research field, at an NAEd breakfast. The NAEd conversation led us to propose a symposium on youth citizenship for the annual NAEd meeting in Washington. Ed Haertel, that year’s program chair, recommended the symposium to Executive Director Gregory White and then President Susan Fuhrman, who not only accepted the idea but also decided to confer the theme of “Beyond the 3R’s” on the meeting.

This was almost by definition an endorsement and actual instantiation of a broad vision by leaders of one of education's most esteemed institutions.

Needless to say, this opportunity provided us with welcome encouragement, as well as a chance to refine our ideas regarding improved ways of approaching youth civic development and education. Both the encouragement and the discussion opportunity were invaluable: working in this marginalized area, as one of our speakers pointed out, often had felt like forays into a rarely visited wilderness.

Jim Banks and I followed up the NAEd symposium with a major conference at Stanford with speakers such as Diana Hess, Carol Hahn, Eric Liu, Marcelo Suarez-Orozco, and Frederick Hess. This splendid group produced a consensus conference report that we published at Stanford and University of Washington. The report offered specific recommendations, methods, and other guidelines for civic education in schools today. Although there is no room for this report here, it can be found at <https://coa.stanford.edu/sites/default/files/Civic%20Education%20report.pdf>.

The report made its way into the hands of a new California Task Force on reforming civic education (among other places); and I was appointed as an advisory member to the Task Force. In the early days of this Task Force's wake, there has been some actual progress in implementing constructive reforms to the state's approach to civic education in its public schools. This may be a part of an upsurge of concern and related activity nationwide on this urgent matter. It is always difficult, of course, to gauge such activity (or its eventual effect) while it is happening; but there is no harm in being optimistic about any movement that one desires. As for the effect of our own NAEd-initiated efforts on this long hoped-for movement, I assess this by my perennial grain-of-sand-on-the-pile philosophy: every contribution, however small, helps in ways that may make a critical difference eventually.

And so it was with the NAEd's role in this small story. By providing an association where its members could work up new ideas, by providing a welcoming forum they could present the new ideas to an assembly of its other members, and by offering encouragement for ideas that departed significantly (and intentionally) from mainstream public policy priorities of our field, the NAEd put its grain of sand on the side of the pile favoring broader visions for American education. I was grateful that the leadership of the NAEd made that choice, and I hope that the NAEd continues to make similar convention-challenging choices throughout its consequential future.

References

- Dillon, S. (2011, May 4). Failing grades on civics exams called a "crisis." *The New York Times*, p. A23.
- Duckworth, A. & Seligman, M. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16, 12, 939–944.
- Gould, J. (Ed.). (2011). *Guardian of democracy: The civic mission of schools*. Philadelphia, PA: The Leonore Annenberg Institute for Civics of the Annenberg Public Policy Center at the University of Pennsylvania. Retrieved from <http://www.civicmissionofschools.org/site/guardianofdemocracy>.
- National Center for Education Statistics. (2011). *The nation's report card: Civics 2010*. (NCES 2011-466). Washington, DC: Institute of Education Sciences, U.S. Department of Education. Retrieved from <http://nces.ed.gov/nationsreportcard/pdf/main2010/2011466.pdf>.
- Ravitch, D. (2010). *The death and life of the great American school system: How testing and choice are undermining education*. New York: Basic Books.
- Schmitt, G. J., Hess, F. M., Farkas, S., Duffett, A. M., Miller, C., & Schuette, J. M. (2010). *High schools, civics, and citizenship: What social studies teachers think and do*. Washington, DC: American Enterprise Institute.
- Southern Poverty Law Center. (2011). *Teaching the movement: The state of civil rights education in the United States*. Retrieved from http://www.tolerance.org/sites/default/files/general/Teaching_the_Movement.pdf.
- Tough, P. (2012). *How children succeed: Grit, curiosity, and the hidden power of character*. New York: Mariner Books.

Reflections Upon 50 Years of U.S. Research in Early Childhood, 1962–2012: A Role for the National Academy of Education

*Diana Slaughter Kotzin*¹

The National Academy of Education can do more to publicize early educational research known to benefit young children and prepare them for the rigors of later schooling. Since the 2008 recession in the United States, increasing attempts have been made to balance budgetary shortfalls by slowly and systematically reversing commitment to one of the few initiatives that has proven beneficial to the long-term educational development and learning of urban children: early childhood education and intervention.

For example, consider the role of play in children's early learning and development since the 2001 emphasis on basic skill acquisition as advanced in the legislation, commonly referenced as "No Child Left Behind." Even before this recession, contextual and social concerns in urban communities were linked to a reduction in actual opportunities for urban children to engage in any form of play behavior (e.g., elimination of recess in many city schools) stressing linear development of skills, possibly at a cost of attention to elaboration of children's conceptual and imaginative abilities. Structured early interventions designed to enhance children's play have definitely been minimized.

Nonetheless, however important such linear academic skills are to school success, it is also important to reaffirm a research tradition that

¹ Diana Slaughter Kotzin is the Constance E. Clayton Professor Emerita in Urban Education at the University of Pennsylvania. She was elected to the National Academy of Education in 2012.

emphasizes the significance of varied approaches to developing literary and social competence during the early childhood years, preschool through grade 4. Close ties exist between social and pretend play behaviors and children's emergent linguistic and social competence. In addition, the critical importance of the arts (broadly defined as pictorial and visual, theatrical, and musical) means that it should be incorporated in urban children's academic learning and development. Scientific research today clearly points to the important contribution of play and the arts to the learning and development of all children, regardless of socioeconomic status, race, or ethnicity. Early childhood intervention has consistently enhanced the cognitive, social, and emotional development of young children, regardless of economic and cultural backgrounds, with long-term evidence of benefits even into adulthood.

Furthermore, parental involvement in children's early learning and development has consistently been linked to children's educability. Public awareness in the future could be raised if the Academy were to highlight the ties between education and schooling *and* early learning, development, and parental education and involvement. We are rapidly becoming a nation with children whose families expect all the "heavy lifting" (i.e., educating) to be done by schools. As a prestigious and generally wise institution, the Academy knows better and could effectively counter these and other public misunderstandings.

Systemic Problems Require Systemic Solutions

Marshall S. Smith^{1,2}

This is a story of two times. Twenty-five years ago, Jennifer O'Day and I argued for systemic reform driven by high-quality standards to address the overall issues of education quality and equity. Times have changed. Today, we continue to support high-quality standards and aligned resources, but we expand the necessary components to include the application of continuous improvement throughout the system, in-school targeted interventions, and partnerships between schools and other institutions that impact the development of children. We also propose a multi-faceted strategy for change that balances support and pressure in productive ways. The paper builds on the experience of the past 2.5 decades. The time is now ripe to rethink the framework of our education system.

Twenty-five years ago—and 25 years since the creation of the National Academy of Education and Title I—Jennifer O'Day and I wrote a paper that set out a new vision for U.S. education (Smith & O'Day, 1991). In his recent book, Jack Jennings (2015, p. 68) referred to the paper as “the intellectual framework” for the reforms of the 1990s. After framing an argument that the existing system in the late 1980s lacked coherence and attention to higher order skills, we argued for “systemic reform” that

¹ Marshall S. Smith is a Visiting Scholar at The Carnegie Foundation for the Advancement of Teaching. He was elected to the National Academy of Education in 1989.

² I wish to thank Jennifer O'Day for her years of collaboration. This article is largely based on our joint scholarship.

would include organizing and aligning the fundamental components of teaching and learning, teacher training, professional development, curriculum, and assessment into a strategy driven by high-quality and challenging content and performance standards. In a second paper written 2 years later, we focused on issues of equality and extended the argument for systemic reform to include what we called “delivery standards” and was later termed “opportunity to learn standards” (O’Day & Smith, 1993).

Both papers were written with the contextual backdrop of the Reagan administration and the early part of the George H. W. Bush administration. At the end of the Reagan years, the 1988 Elementary and Secondary Education Act (ESEA) amendments passed and contained provisions requiring state testing and accountability in mathematics and English in three grades for Title I schools but did not require the alignment of teaching and learning resources and activities to content and performance standards. The Bush administration proposed legislation that included voluntary national standards; however, this legislation failed. Meanwhile Diane Ravitch, then the assistant secretary for the Office of Educational Research and Improvement (OERI), held competitions for the development of voluntary national standards in different subject-matter areas.

A congressionally established body, the National Council for Education Standards and Assessment, also proposed *national* standards and assessments. However, as the ideas about standards stewed in the Congress and in other places, the 10th Amendment—reserving all powers not enunciated in the Constitution for the states—showed its muscle and national standards and assessments were replaced in policymakers’ minds by the approach of giving each of the states control over the nature and content of their own standards and assessments.

The Clinton administration also envisioned standards-based reform. By the fall of 1993 the administration had drafted two bills, Goals 2000 and the reauthorization of ESEA including a new Title I. Goals 2000 was a voluntary grant program to the states to help them implement a standards-based system. Title I of ESEA included language that required all Title I schools to implement standards-based reforms including a testing and accountability regime similar to the one in the Bush administration’s 1988 amendments. After some debate and changes each bill passed the Congress and was signed into law in the spring of 1994.

The idea of Title I requiring standards and assessments to structure teaching in America’s Title I schools was groundbreaking. However, the Title I bill contained another provision that went substantially further than past federal legislation. It required Title I schools to implement standards-based reforms in order to receive Title I monies, and it also

required that the standards for the Title I schools and students were *the same standards as are used in the other schools and students in the state*.³

This was extraordinary. It meant that acceptance of Title I money expanded the reach of Title I directives from low-income schools within a state to a directive that applied to the entire state education system. All students, teachers, schools, and districts would have to conform to a single set of standards, assessments, and accountability. Instead of legislation focused only on a particular group (such as students in low-income schools, students in special education, or bilingual students) or a particular intervention such as professional development, which was the typical approach of federal legislation, Title I of the ESEA 1994 amendments directly affected everyone in the state system. The threat of not receiving Title I funds was a lever to dramatically change the state education systems. This is viewed by Jack Jennings (2015, p. 71) as a major departure from “a centuries-old tradition of local control.”

Although the bills passed with bipartisan support, they did so without the requirement of opportunity to learn (OTL) standards. OTL standards were seen as too great of an intrusion—the 14th Amendment kneeled to the 10th Amendment. This was a substantial loss—the OTL standards could have been a lever to encourage the states and districts to provide resources and support to low-income and minority students in ill-equipped schools.

The Clinton administration expected a reasonable level of implementation of the reforms in all 50 states by 1998, 4 years after passage. We were naïve. It took over a year for federal regulations to be drafted and implemented, some states had to amend their constitutions, and almost all states had to create new legislation to effect implementation. We found that we had totally underestimated the time to fully implement a reform of this complexity. In 2000 *Education Week* found that implementation was at least well under way in many states.

After George W. Bush was inaugurated, waiting for the reforms to take hold was not enough for the new administration. Additionally, many civil rights groups wanted to quickly see evidence that the reforms were closing the achievement gaps. The pressure increased for changes, particularly in the areas of increased testing and more stringent accountability, which many viewed, then and now, as the primary drivers of reform.

In very early 2002, President George W. Bush signed into law the No Child Left Behind Act (NCLB), the latest version of the ESEA amendments. NCLB, while leaving in place the basic structure of standards-based reforms, greatly increased the degree of testing from three grades

³ See Part A Improving Basic Programs: Subpart 1(b)(1) of the 1994 amendments of ESEA at <https://www2.ed.gov/legislation/ESEA/sec1111.html>.

to seven and established aggressive accountability sanctions for not meeting yearly goals. These new provisions came with a new cycle of implementation—the legislation did not require the states to have all seven of the grade-level assessments in place until 2005–2006.⁴

An irony is that, when using the National Assessment of Educational Progress (NAEP) data as a measure of school improvement to compare the gains in scores from 1996 to 2005–2007 (from Clinton reforms to NCLB) and from 2005–2007 to 2013 (from NCLB to now), the earlier period shows a greater rate of growth than the later in fourth-grade reading and mathematics and eighth-grade mathematics. The gains are small in eighth-grade reading and are slightly greater in the NCLB years. The gaps between White and African Americans and Hispanic Americans closed substantially, particularly in mathematics, although for three of the comparisons the rate of closing has declined in the past few years. Many factors affect the levels of student achievement, and causality is difficult to assess, but at the least these data indicate that a more intense focus on accountability is not associated with a faster rate of achievement growth or gap closing. Indeed the opposite hypothesis might be more accurate. Moreover, some have argued that the intense accountability system has had serious and damaging side effects on teacher morale as well as other factors. Perhaps it would have been better if the Congress had just left the ESEA alone in 2001.

Since President Obama took office, several attempts have been made to amend ESEA (NCLB), but the lack of cooperation in Congress has been a huge barrier. Although ESEA still has not been legislatively modified, the administration has used competitions and waivers to reduce many of its accountability stingers. Secretary Duncan—using less than 1 percent of the total budget of 1 year of K–12 education in the United States—generated a blizzard of activity at the state level using competitions as his lever. In *Race to the Top* a relatively small number of states were awarded grants based on the quality of their proposals and their commitment to adopt a particular set of reforms that were promoted by the administration. However a large number of states competed, many of them making strong commitments to implement the reforms, which they carried out even without the funding.

The competitions included a piecemeal mixture of reforms that had credibility in the “reform” community including holding teachers accountable for student achievement as measured by state tests, strategies to improve very low achieving schools, efforts to have states greatly improve their education data system, and adoption of college and career

⁴ See the NCLB section on Academic Standards, Academic Assessments, and Accountability at <https://www.govtrack.us/congress/bills/107/hr1>.

ready standards. The Secretary continued his reform efforts by predicating waivers of NCLB accountability provisions with requirements that states adopt these reforms if they want to obtain the waivers.

Although both of Duncan's reform strategies had been preceded by an even more extensive example of leveraging in the 1994 amendments his efforts have been roundly criticized as federal over-reach. Moreover some evidence suggests that the teacher accountability reforms based on test results together with the testing regime of NCLB have had few positive effects and some substantial negative effects that include reduced teacher morale, reduced retention, and fewer potential teachers entering pre-service training. The attempt to substantially improve the lowest scoring schools also lacks clear evidence of success—this is a very substantial problem that deserves great attention. Although the emphasis on states to adopt college and career standards also stimulated criticism of federal over-reach, it does not seem to have greatly harmed the health of the important Common Core standards. In my view the efforts to upgrade the data systems have been successful and important.

Thus, the sum of these recent reforms of the Obama administration, I think, is slightly on the positive side of the ledger. However, we can take away two important lessons on the negative side. The task of creating effective and fair teacher accountability systems is far more complex than the Department realized, and the Secretary's efforts may have stimulated negative by-products—this is not dissimilar to the kind of reaction that resulted from the implementation of NCLB's very intense accountability strategy. Unless you know deeply what you are doing when strong interventions are carried out Newton's third law will create havoc. Finally the failed struggle to improve very low scoring schools even with quite a bit of money should act as a stimulant to all of us to not only do better but also realize that schools often cannot go it alone—"it takes a district" and often engagement of the local and the professional communities, as well.

Up to here most of this short history is from a federal perspective, which, in the overall picture is quite narrow. A lot has happened and been learned since 1990 from both practice and research in states, districts, and schools. We have learned that serious and successful continuous improvement takes a long time and benefits from a clear vision and sustained, smart, and adaptive collegial effort across and entire system. We also know that tailored data systems can be of great support to improvement efforts. We have rediscovered one of Jack Carroll's components of school learning and Sheldon White's passion: the importance of motivation. Our understanding of the power of other components of Social Emotional Learning (SEL) has also flourished including Jim Coleman's ideas about the importance of a student's "control over environment," which he set out in the Equality of Educational Opportunity Report in 1966, 49 years

ago. Theories and examples of approaches to continuous improvement and improvement science in the business and healthcare worlds have been around for years but it is only in the past two decades that they have been seen as critical in education (Bryk, Gomez, Grunow, & LeMahieu, 2015). We have learned about the power of networking and professional communities and their effect on the quality of experiences of professionals. We now more deeply understand how to teach in a wide variety of areas and how to engage students in productive discourse, the critical importance of early intervention, and even ways of using technology to substantially enhance learning. In addition, we have learned a great deal about the effects of poverty and community on the brain, mindset, library of experiences that students draw from, and opportunities that they have out of school.

From our experiences in the areas of school reform we have learned that the value of separate interventions or “magic bullets” is limited. Over the past 25 years vouchers have been tried and failed in our system many times, charter schools have had limited success in particular areas but are not game changers, intense accountability has serious side effects, and isolated technology interventions are often distracting and not particularly useful.

In some ways, however, 2015 is quite similar to 1990. We still appear to be in awe of some other nations’ successes in education, regardless of how different they are from the United States. We face a similar though greater problem related to income inequities and sometimes fool ourselves into thinking that schools alone can solve the academic gaps and even the income inequalities. We recognize the negative effects on students of poverty at home and neighborhoods with high densities of poverty, but we do not seem to do much about them. We tolerate huge disparities in human and capital resources among schools within districts, districts within states, and among states. We seem to have accepted standards but we still argue about a common or national set of standards. We continue to greatly under estimate the importance of implementation. Finally apart from a modest consensus about the importance of the common core with its pedagogical implications and formative assessment we seem to lack a unified vision of what our education systems should look like.

Although where we are now in 2015 is far more complex than I can describe in this short note, the issues I have noted create the backdrop for a forthcoming article prepared by Jennifer O’Day and me (O’Day & Smith, 2015). What follows is a very brief sketch of this new paper—the article has rich documentation and examples and far more thoughtful arguments than I have included in the remainder of this note.

The article was influenced by what Jennifer and I have learned from scholarship and from the struggles of the U.S. education system over the

past 25 years. We start with the tenet that we will never achieve equality of opportunity or outcomes in the system unless we have quality systems, and we will never achieve quality if we have inequality. A second tenet is that the classrooms, schools, districts, and states are complex environments—rigid requirements are often insensitive to important differences in the environments. Flexibility and adaptability are critical if we expect to achieve high-quality and effective systems. A related tenet is that we need a productive balance of support and pressure—recently the federal government has tilted much too strongly to the pressure side of the equation. We need to rebalance, and no single source of support and pressure will be sufficient. Improvement and support should trump top-down accountability. A fourth tenet is that while our rhetoric has often been strong, on balance our actions in the areas of civil rights required by the 14th Amendment, Title IX, Lau, and the Education for All Handicapped Act have been weak in the past 25 years. This tenet along with the need for adaptability suggests the need for a substantially increased level of federal intervention in the civil rights areas with a greatly reduced role in the day-to-day instructional, curriculum and accountability components of schools. A final tenet is that thoughtful and sustained implementation is necessary for the success of any complex intervention.

In the new article, while acknowledging the unequal opportunities outside schools that contribute to the disparities in educational achievement, attainment, and various indicators of adult success, we focus on addressing inequities within K–12 education. We argue that disparities within the educational system are the product of institutional structures and cultures that both disenfranchise certain groups of students and depress quality overall. Systemic causes require systemic solutions, and we envision a three-pronged systemic remedy: a high-quality, standards-based, instructional system and a continuous improvement approach for addressing the quality of educational opportunities for underserved students as well as of the system as a whole; targeted high-leverage interventions consistent with the overall approach but focused on SEL, language, response to intervention and key transition points and needs; and stronger connections between schools and other institutions and systems affecting the development and well-being of children and youth.

We then set out a detailed and aggressive change strategy that incorporates both pressure and support for improvement from three distinct but interacting sources: government and administrative policy (federal, state, and local); professional accountability and networking; and collective engagement of parental, community, and advocacy organizations.

We recognize that this reform cannot be legislated through voluntary federal grants and the leverage of Title I as were the reforms in the 1990s. It will not come about as a top-down edict. It will have to bubble

up from the schools, districts, and states. Finally, it will be a while until we see the reform as a unified whole. One path that might be taken is that districts and states will move relatively slowly as first the Common Core is implemented and then the effort and results of this implementation stimulate the beginnings of overall continuous improvement as they learn that effective interventions require an ability of the organization to steadily focus and support the learning of all of the students. There are other paths. However for all of them the quality and support of the political and local environments are critical for success.

References

- Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.
- Jennings, J. (2015). *Presidents, Congress, and the public schools: The politics of education reform*. Cambridge, MA: Harvard Education Press.
- O'Day, J. A., & Smith, M. S. (1993). Systemic school reform and educational opportunity. In S. Fuhrman (Ed.), *Designing coherent education policy: Improving the system*. San Francisco, CA: Jossey-Bass.
- O'Day, J., & Smith, M. S. (2015). Quality and equality in American education: Systemic problems, systemic solutions. To be published in a book on Inequalities by ETS in the Fall.
- Smith, M. S., & O'Day, J. A. (1991). Systemic school reform. In S. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing* (pp. 233–267). London, UK: Falmer Press.

50 Years of Research on Reading

Catherine Snow¹

The 1960s, when the National Academy of Education (NAEd) was established, were a time of spirited debates about the best approaches to literacy education. Some argued for an early, intensive, and systematic focus on teaching the code as a code, to help children navigate the complexities with which sounds are mapped to letters in English. Others argued for a focus on meaning, with the assumption that children could figure out the mapping with incidental help from adults if they understood the larger message of the text. This controversy, together with the unflinching conviction of correctness on the part of combatants on both sides, was fed by some basic facts: many children can learn to read no matter what instructional approach is adopted, and most primary teachers were using an eclectic approach in any case.

Nonetheless the controversy raged. Its virulence no doubt did a disservice to educational practitioners seeking instructional guidance, as well as to educational researchers seeking credibility (as documented, e.g., by Carl Kaestle in his report *The Awful Reputation of Educational Research*, 1992). NAEd member Jeanne Chall, never one to duck a fight, faced the controversy head-on, with her book *Learning to Read: The Great Debate* (1967). That book reviewed and sifted the evidence about reading instruction. Chall never denied the possibility that children could learn to read from meaning-focused approaches, or from only modestly systematic

¹ Catherine Snow is the Patricia Albjerg Graham Professor at the Harvard Graduate School of Education. She was elected to the National Academy of Education in 1996.

methods of teaching them about the code. But she was particularly concerned about children growing up in poverty—children who were most dependent on school for literacy achievement. She argued strongly that such children were disadvantaged by approaches to reading instruction that under-emphasized phonics.

Chall served the NAEd loyally in its efforts to improve the relevance and utility of literacy research to national challenges of upgrading literacy achievement. She co-chaired with John Carroll an NAEd report titled *Toward a Literate Society* (National Academy of Education, 1975), a response to a request from the assistant secretary of the U.S. Department of Health, Education and Welfare and the U.S. Commissioner of Education for guidance about implementing the Right-to-Read program, launched in 1969. It is instructive to compare the “comprehensive review” of knowledge about literacy produced by the Carroll-Chall chaired Commission on Reading to the 1985 report produced by a subsequent NAEd Commission on Reading, *Becoming a Nation of Readers* (Anderson, Hiebert, Scott, & Wilkinson, 1985). The Commission that produced the 1985 report was chaired by NAEd member Richard Anderson and included other current and future NAEd talent. Although the quantity of research available for synthesis in the 1985 report was much greater, the conclusions were not radically different, and indeed emphasized both the centrality and the insufficiency of systematic, early phonics instruction.

One might think that two prestigious NAEd reports converging on a shared set of recommendations would have sufficed to end the debate about the best approach to reading instruction. The debate persisted, though, fed by a number of factors:

- As noted above, many children can learn to read with any approach to instruction. (Indeed a small number of children growing up in literate contexts learn to read with no instruction at all.) Thus, many practitioners were encountering success with methods that had no research support. Such practitioners are unlikely to be convinced that alternative methods are better.
- The reading process is sufficiently complex and multiply determined that one can find research support for the value of many different instructional emphases, for at least some subgroups of students.
- The interpretation of the various instructional recommendations, especially those targeting systematic phonics instruction, by textbook publishers and deliverers of professional development were often unbalanced and rigid, leading to backlash against the claims rather than against their implementation (see Snow & Juel, 2005, for a historical review of reading curricula).

An additional reason for the failure to resolve the “great debate,” though, was anticipated by Jeanne Chall in her afterword to *Becoming a Nation of Readers*. She expressed agreement with the general thrust of the report and its recommendations, but noted that the challenge of ensuring universal literacy required special attention, and perhaps more and different treatment, for those with “serious reading difficulties.” Chall went on to characterize learners with reading difficulties very broadly: “children from low income families, ethnic minorities, non-English or recent speakers of English, and those with specific reading and learning disabilities.” That category of struggling readers came into central focus after the adoption of No Child Left Behind of 2001 (NCLB), with its emphasis on ensuring that all children, including those in high-risk groups, be successful in learning to read (and do math). This shift in national focus from instruction for the average child to instruction for the children most at risk re-energized those insisting on systematic and intensive phonics instruction as the key to literacy success.

The elevated profile of the at-risk reader in influencing beliefs about reading instruction was exacerbated during the implementation of the Bush-era Reading First models (U.S. Department of Education, 2002) by the insistence on experimental support for the effectiveness of instructional approaches. The easiest group with which to conduct experimental evaluations of literacy approaches is children who have failed to learn to read; they are identified, classified, and available for random assignment to targeted interventions in small groups, often with specially trained teachers. These experimental studies with delayed or disabled readers showed the value of explicit training in phonological awareness and systematic, intensive phonics, and the results were then generalized to all children. Such findings were elevated to U.S. Department of Education policy via, for example, the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD)-sponsored Report of the National Reading Panel (a panel which, incidentally, included no NAEd members). The National Reading Panel report (2000) was taken as strong guidance by the U.S. Department of Education during the first several years of NCLB, and was to a large extent responsible for a re-emergence of worries about appropriate balance between code and meaning in early reading instruction. It thus displaced to some significant extent the influence of *Preventing Reading Difficulties in Young Children*, a 1998 National Research Council report that incorporated a richer selection of developmental, instructional, and qualitative data as well as a good dose of theory (National Research Council, 1998). *Preventing Reading Difficulties* was initially received with considerable enthusiasm by practitioners and policymakers at both federal and state levels, but was critiqued as being insufficiently “research-based” by some enthusiasts for

experimental rigor. Nonetheless, it did form one strong basis for a 2005 NAEd report titled *Knowledge to Support the Teaching of Reading* (Snow, Griffin, Burns, & the NAEd Subcommittee on Teaching Reading, 2005), part of a larger NAEd effort to systematize the knowledge to which teachers should have access.

The issue of the correct role for experimental findings in determining educational practice is one that has received attention within various fields, not just reading. Of course experiments are the strongest basis for causal inferences about the effectiveness of a treatment, but the correct use of experimental findings requires some caution in extrapolation beyond the group for which effectiveness is demonstrated. The dominance in results about early reading interventions of findings based on troubled and/or delayed readers, combined with an extremely naïve view promulgated early in the Reading First era that most reading difficulties stemmed from students' inadequate grasp of the alphabetic principle, led to the proliferation of phonics-heavy approaches and relative neglect of oral language, vocabulary, knowledge, and meaning-making in early literacy instruction. The medical equivalent would be putting everyone on a gluten-free diet because avoiding gluten makes folks with celiac disease feel so much better.

Interestingly, it has not been literacy research, prestigious panel reports, or actions by members of the NAEd that have expanded attention in literacy instruction to matters that go beyond phonics-based approaches. Rather, the focus on 21st century skills, and the associated development of College-and-Career-Ready standards linked to literacy standards for preK through 12th grade, have replaced "reading correctly and fluently" with "learning, analyzing, critiquing, and synthesizing while reading" as key literacy skills. While such 21st-century skills are indeed dependent on reading correctly and fluently, and thus are possible only if students have had the opportunity to grasp and master the alphabetic principle, they go far beyond the alphabetic principle. The simple assumptions of the early 2000s, that solid first through third grade instruction leading to excellent performance on measures such as the Test of One Word Reading Fluency and the DIBELs would raise the reading achievements of U.S. students to international standards, are shattered. It is clear now that massively more is needed.

Thus, the reading research agenda for the next period of time reunites reading with its traditional goals: reading to learn science, math, and history as well as to analyze literature; reading to understand how others with different experiences understand the world; reading to consider and to critique or to adopt alternative political, religious, or cultural stances; and reading for pleasure, for inspiration, for information, and for the opportunity to share experiences with other readers. The next

NAEd Commission on Reading, if there is another one, will be faced with developing a research agenda that could support teachers, curriculum developers, and test developers in incorporating this full range of reading skills into their practices and products. With any luck, it will not have to start by resolving conflicts between those supporting code-based versus meaning-based approaches to instruction.

References

- Anderson, R. C., Hiebert, E. H., Scott, J. A., & Wilkinson, I. A. G. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington, DC: National Academy of Education. Retrieved from <http://files.eric.ed.gov/fulltext/ED253865.pdf>.
- Chall, J. (1967). *Learning to read: The great debate (an inquiry into the science, art, and ideology of old and new methods of teaching children to read, 1910–1965)*. New York: McGraw-Hill.
- Kaestle, C. (1993). The awful reputation of education research. *Educational Researcher*, 22(1), 23–26–31.
- National Academy of Education. (1975). *Toward a literate society: The report of the Committee on Reading of the National Academy of Education with a series of papers commissioned by the committee*. New York: McGraw-Hill.
- National Academy of Sciences. (2000). *Prevention of reading difficulties in young children*. Washington, DC: National Academy Press.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved from <https://www.nichd.nih.gov/publications/pubs/nrp/Pages/smallbook>.
- National Research Council. (1998). Preventing Reading Difficulties in Young Children. C. E. Snow, M. Susan Burns, & P. Griffin (Eds.), Committee on the Prevention of Reading Difficulties in Young Children, Board on Behavioral, Cognitive, and Sensory Sciences, and Division of Behavioral and Social Sciences and Education.
- Snow, C. E., & Juel, C. (2005). Teaching children to read: What do we know about how to do it? In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 501–520). London, UK: Blackwell.
- Snow, C. E., Griffin, P., Burns, M. S., and the NAE Subcommittee on Teaching Reading. (2005). *Knowledge to support the teaching of reading: Preparing teachers for a changing world*. San Francisco, CA: Jossey-Bass.
- U.S. Department of Education, Office of Elementary and Secondary Education. (2002). *Guidance for the Reading First Program*. Retrieved from <https://www2.ed.gov/programs/readingfirst/guidance.pdf>.

The Importance of Fostering Productive Student Participation

Noreen M. Webb¹

Membership in the National Academy of Education brings the exciting prospect of sharing insights and ideas about thorny educational problems with others who have a great deal of wisdom and experience. An issue that colleagues at the University of California and I have been grappling with is how to engage students as active and confident participants in classroom life—especially in large and small group discussions with their teacher and peers—so that they not only learn, but also develop positive attitudes about their competence and potential.

Several decades of research links student participation to student outcomes, including development of subject-matter knowledge and understanding, reasoning and argumentation skills, ability to converse with others, self-confidence and perceptions of themselves as capable learners and contributors, and willingness to interact with others who may seem different from themselves. Among other things, participating in classroom dialogue helps students learn how to explain their own thinking and listen carefully to others' ideas, and how to elaborate on, as well as challenge, others' suggestions in productive ways. When students try to articulate their thoughts clearly, compare their ideas and perspectives with those offered by others, and explore the ideas that emerge, they may monitor their own and others' thinking, correct misconceptions, fill in knowledge gaps, and develop new ideas and perspectives.

¹ Noreen M. Webb is a Professor at University of California, Los Angeles. She was elected to the National Academy of Education in 2015.

In addition to detailing these benefits for students, the 2015 American Educational Research Association (AERA) volume *Socializing Intelligence Through Academic Talk and Dialogue* documents many ways in which teachers may orchestrate productive classroom dialogue (Resnick, Asterhan, & Clarke, 2015). For example, teachers can establish norms for desired communication and respectful engagement, set ground rules and guidelines for productive participation and individual accountability, provide training and practice in communication skills, and encourage students to elaborate on their own thinking and the ideas of others.

Recently we worked with a small group of elementary school teachers who showed great skill and dedication to providing support for students to engage in deep ways with each other's mathematics ideas. These teachers drew upon a wide repertoire of in-the-moment moves to help students clarify, question, and elaborate their own and others' ideas, recognize connections between ideas, and work together to forge new perspectives and problem-solving strategies. They worked hard to position students as valued and capable contributors (Franke, Ing, Johnson, Turrou, & Webb, in press). Many interactions among students in these classrooms, as well as between students and teacher, were highly productive, with students showing sustained, synchronous, and deep engagement with each other's ideas, and collaborating in ways that led to new problem-solving approaches and mathematical understanding.

However, not all interactions were that productive. Sometimes students could not figure out how to make their ideas heard and understood, or could not make sense of a partner's idea, or did not leave space for their partners to share their thinking. Certain students encountered these situations repeatedly.

These challenging interactions may have consequences for student learning, their perceptions about their ability to do the work and to contribute to conversations, and their attitudes about working with others. Our outcome data show this. While many students said they liked to share ideas and work with others, and were confident in their abilities, exceptions remained. Some students admitted that they did not like to share their ideas, sometimes saying that showing their work to classmates made them scared, nervous, or sad. Other students said that they were willing to share their strategy only when they were confident that their answer was correct; but even then, a few were apprehensive that their classmates might disagree, or point out errors. Some students expressed frustration at working with classmates, especially when they did not have a chance to share their ideas, or felt disrespected or ignored. Some students came away saying they were not good in math. Yet such self-doubts did not always match evidence about their achievement; some of these

students showed high scores on statewide standardized tests, teacher-designed tests, and our researcher-designed assessment.

The approach to promoting positive student outcomes by supporting students' active participation and engagement in each other's ideas is highly promising, and supportive evidence is steadily accumulating. However, even in the exceptional classrooms we observed, interactions occurred that were not as productive socially or mathematically as one might want them to be, and some students struggled to figure out how to participate. Although we do not expect that students should always engage seamlessly with each other's ideas, a challenge is this: How can we help teachers create classroom environments in which interactions are generally productive for each and every student? Addressing this challenge will take the collective effort of experienced, thoughtful, and wise minds representing multiple perspectives and fields.

References

- Franke, M. L., Ing, M., Johnson, N. C., Turrou, A. C., & Webb, N. M. (in press). Dude, why did you start without me? Fostering engagement with others' mathematical ideas. In C. Chinn, E. Manalo, & Y. Uesaka (Eds.), *Promoting spontaneous use of learning and reasoning strategies: Theory, research, and practice*. New York: Routledge.
- Resnick, L. B., Asterhan, C. S. C., & Clarke, S. N. (Eds.). (2015). *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Education to Promote All Students' Social, Emotional, and Academic Competence

Roger P. Weissberg^{1,2}

I am proud to be a member of the National Academy of Education (NAEd). For 50 years the Academy has advanced rigorous education research and its applications to policy and practice. I have aspired to do the same during my 40-year career—with a focus on advancing research, practice, and policy in the area of social and emotional learning (SEL). SEL involves the processes through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions.

Imagine, if you will, two different schools. One educates children who do well on tests of academic achievement. Another educates children who do well on academic tasks and are also socially and emotionally competent, children who are on the path to becoming knowledgeable, responsible, capable, and caring adults who possess the essential tools to succeed in college, career, and life. Which school would you want all American students to attend?

¹ Roger P. Weissberg is the NoVo Foundation Endowed Chair in Social and Emotional Learning at University of Illinois at Chicago. He was elected to the National Academy of Education in 2013.

² Work on this essay was supported, in part, by grants from the Einhorn Family Charitable Trust, NoVo Foundation, and 1440 Foundation. I also appreciate the efforts of Teresa Borowski, Karen Niemi, Hank Resnik, and Jennifer Schneider for their thought partnership in preparing this manuscript.

Over the years my collaborators and I have conducted extensive research that has laid the foundation to transform the educational system of America to look like the second school. Creating schools where children are prepared for the challenges of the world as positive family members and neighbors, contributing citizens, and productive workers is not just wishful thinking. There is growing evidence that implementing systemic, evidence-based SEL will help to achieve this reality (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011; Weissberg, Durlak, Domitrovich, & Gullotta, 2015).

The Collaborative for Academic, Social, and Emotional Learning (CASEL) introduced and described the field of SEL in the groundbreaking book *Promoting Social and Emotional Learning: Guidelines for Educators* (Elias et al., 1997). CASEL categorizes the knowledge, skills, and attitudes that play a role in social and emotional competence into five domains: self-awareness, self-management, social awareness, relationship skills, and responsible decision-making (Zins, Weissberg, Wang, & Walberg, 2004). SEL involves fostering social and emotional competencies, which can be taught and learned, through instruction in which social and emotional skills are taught, modeled, practiced, and applied in various situations. SEL programming is based on the understanding that the best learning occurs in the context of supportive relationships that make learning challenging, engaging, and meaningful. It enhances both social and emotional competence and academic performance by creating school climates that are safe, caring, and cooperative. Educators, practitioners, researchers, and policymakers can use SEL as a conceptual framework to promote the social, emotional, and academic competence of young people and coordinate school, family, and community programs and practices (Durlak, Domitrovich, Weissberg, & Gullotta, 2015; Weissberg & Cascarino, 2013).

The importance of interpersonal and intrapersonal competencies can be seen through the growth of kindred fields such as noncognitive skills, 21st-century skills, character education, deeper learning, and the whole child approach (National Research Council, 2012). In addition, teachers and employers agree that fostering social-emotional skills and positive mindsets through real-life learning opportunities where students apply these competencies is a powerful strategy to improve academic performance and student development. A national survey of teachers found that 93 percent want more focus on SEL in schools (Bridgeland, Bruce, & Hariharan, 2013). Employers echo this sentiment, saying teamwork, problem solving, and good character are qualities they seek in employees entering the workforce (Ee & Chang, 2015).

Rigorous scientific research has produced convincing evidence that high-quality, well-implemented SEL can result in multiple positive academic, social, and emotional outcomes (Durlak et al., 2011; Sklad,

Diekstra, De Ritter, Ben, & Gravesteyn, 2012). These major meta-analytic reviews found that SEL programs improved social and emotional skills, self-concept, attitudes and connection to school, and classroom climate. Students who received SEL programming also demonstrated more prosocial behavior, less emotional distress, and fewer conduct problems such as disruptive classroom behavior, aggression, bullying, and delinquencies. Children who experienced social and emotional education also did better academically, demonstrating an 11-percentile increase in achievement scores. In addition to these short-term outcomes, greater social and emotional competence produces benefits in the long term such as increased likelihood of high school graduation, readiness for postsecondary education, career success, positive family and work relationships, better mental health, reduced criminal behavior, and engaged citizenship (Hawkins, Kosterman, Catalano, Hill, & Abbott, 2008; Jones, Greenberg, & Crowley, 2015).

Social and emotional competence can be promoted on a large scale. In 2011 CASEL launched a national initiative aimed at supporting school districts' capacities to promote SEL for all students. CASEL's Collaborating Districts Initiative (CDI), which includes 8 of the 200 largest urban school districts in the nation, recognizes that positive student outcomes depend on improving classrooms and schools, which in turn depends on improving districtwide capacities and conditions. The initial results of an extensive evaluation of CASEL's CDI show that it is possible to develop districts' capacities to plan, implement, and monitor systemic changes that impact schools and classrooms in ways that enhance students' social-emotional development and academic performance.

Clearly SEL is achievable and produces important benefits for young people. In addition to the educational benefits, however, SEL has important economic benefits. One recent study demonstrated an \$11 return for every \$1 invested in SEL programming (Belfield et al., 2015).

The increased emphasis on the importance and value of social and emotional competencies presents numerous questions: How balanced is education? What are the goals of education? Are we effectively working to achieve these goals? When we ask people around the country what they want children to know and be able to do when they graduate from high school, parents and teachers emphasize social and emotional skills such as the ability to persist, empathize with others, and manage their behavior so they can achieve challenging goals. Are we making this happen? Is the importance of such skills reflected in our standards and practices?

Considering the powerful combination of evidence and support for SEL, another key question for the field as we move forward is: for those committed to the broader vision of SEL, what agenda needs to be in place so that by 2025 (when the NAEd celebrates its 60th anniversary), SEL will

be established in 50 percent of the nation's schools, with progress being made toward even greater numbers? My CASEL collaborators and I have developed a five-pronged approach to achieve this goal over the next decade.

First, there is a need for a National Commission to build consensus about balanced education in our fragmented field. CASEL and many thought leaders from kindred fields will work together to align and advance research, practice, and policies that enhance the social, emotional, and academic development of students. The range of labels, terminology, inconsistent language, specialties, and uncoordinated interventions related to social and emotional development creates barriers for districts and schools to prioritize social-emotional competence in integrated, systemic ways. By bringing together field leaders, our strengthened collective voice will serve to promote and prioritize social and emotional development on a national scale. It will also help to support widespread implementation of evidence-based SEL practice.

The National Commission will advance three streams of work to lay the proper foundation for the establishment of clear, aligned goals, a plan of action, and a framework for implementation: (1) a *research* stream to broadly assess the quality of research in the field and the findings relevant to students and educators; (2) a *practice* stream to review and identify the best approaches to designing, implementing, and continuously improving systemic preschool to high school curriculum and instruction, as well as to identify the best ways to scale implementation and professional development; and (3) a *policy* stream to assess federal and state education policy and draft recommendations for policymakers.

Second, we need better practical social and emotional competence assessments that are scientifically sound, developmentally appropriate, feasible to administer, affordable, and actionable (Denham, 2015; Marzano, 2015). Because what gets assessed is more likely to get addressed in schools, the development of social and emotional competence assessments for preschool to high school students is one of the most pressing and critical issues for the field to tackle. CASEL is working with nationally renowned researchers and knowledgeable practitioners to identify and establish SEL measurement tools. This work is being conducted through a two-fold, parallel process: (1) an examination of the current state of social and emotional assessment tools in order to provide immediate support for educators in the field by highlighting high-quality measures and best practices for using them and (2) the creation of guidelines for educators, researchers, and developers who want to create the next generation of social and emotional assessments and put them into practice in schools across the nation.

Third, conditions that support nationwide implementation of high-

quality, systemic district- and school-wide SEL need to be documented (Mart, Weissberg, & Kendziora, 2015; Weissberg & Cascarino, 2015; Weissberg, Durlak, Domitrovich, & Gullotta, 2015). Such conditions include guidelines and tools to (1) develop a shared vision for SEL among all stakeholders; (2) assess needs and available resources for district- or school-wide SEL implementation; (3) provide ongoing and embedded professional learning to administrators, teachers, and support staff; (4) integrate evidence-based classroom, school-wide, family, and community SEL programming; and (5) establish systems for measurement and continuous improvement. To this end, CASEL will continue our long-term partnership with the eight CDI districts and experts in the fields of practice, research, policy, and communications to share with educators across the nation a clear framework, evidence, models, tools, and resources to bring SEL into every classroom across the country.

Fourth, CASEL is working with field leaders, practitioners, and policymakers to develop model student learning standards and implementation guidelines and to ensure that federal policy and funding support SEL research and practice (Dusenbury et al., 2015). Learning standards serve as a roadmap for curriculum development and instruction, legitimize the importance of SEL, and help lower barriers to implementation. They also focus professional preparation and development and inform assessments of student learning. Without such standards, school leaders will struggle to build community support to implement SEL programs, define targeted learning goals and outcomes, and find adequate funding for implementation. Significant advances thus far include the fact that all 50 states have adopted preschool social and emotional development standards, while Illinois and a few other states have developed comprehensive SEL standards for preschool through grade 12.

Finally, federal policies regarding SEL can provide schools, districts, and communities with public resources to support their SEL efforts (Zaslow, Mackintosh, Mancoll, & Mandell, 2015). With more policymakers committed to advocating for SEL and more policies in place, we can create the conditions for systemic implementation of SEL in more districts, as well as raise awareness, demand, and support for SEL. CASEL and the National Commission will work to ensure that education legislation recognizes the value of initiatives that promote social and emotional competencies in addition to academic competencies.

Henry Ford said, "Coming together is a beginning, keeping together is progress, working together is success." The National Commission will be our new beginning. Our work in assessment, with school districts, on state standards, and on federal policies over the next several years will motivate great progress. Our collaborative gains will lead to success for the field and the country. Following the mission of the NAEd, we will

continue to move research, policy, and practice in clear directions that will provide all our children with opportunities to succeed in school and life.

References

- Belfield, C., Bowden, B., Klapp, A., Levin, H., Shand, R., & Zander, S. (2015). *The economic value of social and emotional learning*. Center for Benefit-Cost Studies in Education, Teachers College, Columbia University. Retrieved from www.cbcse.org.
- Bridgeland, J., Bruce, M., & Hariharan, A. (2013). *The missing piece: A national survey on how social and emotional learning can empower children and transform schools*. Washington, DC: Civic Enterprises.
- Denham, S. A. (2015). Assessment of SEL in educational contexts. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 285–300). New York: Guilford.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.
- Durlak, J. A., Domitrovich, C. E., Weissberg, R. P., & Gullotta, T. P. (Eds.). (2015). *Handbook of social and emotional learning: Research and practice*. New York: Guilford.
- Dusenbury, L. A., Newman, J. Z., Weissberg, R. P., Goren, P., Domitrovich, C. E., & Mart, A. K. (2015). The case for preschool through high school state learning standards for SEL. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 532–548). New York: Guilford.
- Ee, J., & Chang, A. (2015). *Preparing youths for the workplace*. Hackensack, NJ: World Scientific Publishing Company.
- Elias, M. J., Zins, J. E., Weissberg, R. P., Frey, K. S., Greenberg, M. T., Haynes, N. M., ... & Shriver, T. P. (1997). *Promoting social and emotional learning: Guidelines for educators*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hawkins, J. D., Kosterman, R., Catalano, R. F., Hill, K. G., & Abbott, R. D. (2008). Effects of social development intervention in childhood 15 years later. *Archives of Pediatric Adolescent Medicine*, 162(12), 1133–1141.
- Jones, D., E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, e1–e8. Published online ahead of print. doi: 10.2105/AJPH.2015.302630.
- Mart, A. K., Weissberg, R. P., & Kendziora, K. (2015). Systemic support for SEL in school districts. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 482–499). New York: Guilford.
- Marzano, R. J. (2015). Using formative assessment with SEL skills. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 285–300). New York: Guilford.
- National Research Council. (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. J. W. Pellegrino & M. L. Hilton (Eds.), Committee on Defining Deeper Learning and 21st Century Skills, Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Sklad, M., Diekstra, R., De Ritter, M., Ben, J., & Gravesteyn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, 49(9), 892–909.

- Weissberg, R. P., & Cascarino, J. (2013, October). Academic learning + social-emotional learning = national priority. *Phi Delta Kappan*, 95(2), 8–13.
- Weissberg, R. P., Durlak, J. A., Domitrovich, C. E., & Gullotta, T. P. (2015). Social and emotional learning: Past, present, and future. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 3–19). New York: Guilford.
- Zaslow, M., Mackintosh, B., Mancoll, S., & Mandell, S. (2015). Federal policy initiatives and children's SEL. In J. A. Durlak, C. E. Domitrovich, R. P. Weissberg, & T. P. Gullotta (Eds.), *Handbook of social and emotional learning: Research and practice* (pp. 549–565). New York: Guilford.
- Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (Eds.). (2004). *Building academic success on social and emotional learning: What does the research say?* New York: Teachers College Press.

The Education of a Diverse Learning Population

Income Inequality and American Education

Greg J. Duncan and Richard J. Murnane¹

A century ago, assembly-line production methods made cars and other consumer durables affordable for millions of Americans, in the process generating millions of low-skilled but quite well-paying jobs. Over the past 40 years, computer-driven technological changes have transformed the world's economy while driving up the employment and earnings of highly skilled workers. The widening pay gap between high- and low-skilled workers is responsible for much of the sharply growing gap in total family incomes between high- and low-income families.

Technological changes and rising inequality pose enormous problems for America's public schools and the dream of socioeconomic mobility for low-income families. By upgrading the skills required in hundreds of middle-class occupations, technological change has increased what the nation asks of its schools. For example, 40 years ago, high school graduates who read well enough to follow directions gained access to many jobs paying middle-class wages. Today, requisite literacy skills include the ability to form reasoned judgments about the accuracy and usefulness of

¹ Greg J. Duncan is Distinguished Professor in the School of Education at the University of California, Irvine. He was elected to the National Academy of Education in 2009. Richard J. Murnane is the Thompson Research Professor at the Harvard Graduate School of Education. He was elected to the National Academy of Education in 1990. They are the co-editors of the 2011 book *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, New York: Russell Sage, and co-authors of the 2014 book *Restoring Opportunity: The Crisis of Inequality and the Challenge for American Education*, which was published by the Harvard Education Press and the Russell Sage Foundation.

the hundreds of responses that an Internet search on topics ranging from transportation to health insurance options generates. At the same time, the increase in income inequality has affected where families live and how much money they can spend to nurture their children's capacities. These changes have placed great strains on America's decentralized approach to public education, particularly in schools serving large numbers of children from low-income families.

An obvious advantage of a higher family income is it enables parents to buy books, computers, high-quality child care, summer camps, music lessons, private schooling, and other enrichments for their children. As we document in our edited volume *Whither Opportunity*, in the early 1970s, the richest 20 percent of families spent about \$3,000 more per child per year (in 2012 dollars) on child enrichment than did the poorest 20 percent. By 2006, this gap had nearly tripled, to \$8,000 per child per year. This adds up to a \$100,000 spending gap over the course of a child's primary and secondary schooling—a huge amount. Enrichment experiences matter because they help children acquire the vocabulary and background knowledge critical to achieving the high levels of literacy needed for many kinds of well-paying work.

Less obvious is that as income inequality has increased, so has the residential isolation of low- and high-income families. Relative to 40 years ago, poor families are now more likely to be surrounded by other poor families, while high-income families are similarly isolated. Because most children still attend schools close to their homes, rising residential segregation has led to increasing concentrations of low- and high-income children attending separate schools. The resulting changes in student body composition have shaped school functioning and contributed to the increasing gap between the achievement and completed schooling of children growing up in low- and high-income families.

A child from a poor family is now two to four times as likely as a child from an affluent family to have classmates with behavioral problems and low academic skills. This sorting matters, because the weak cognitive skills and behavioral issues of many low-income children have a negative effect on their classmates' learning. They also make it difficult to manage classrooms in which students are working together in small groups.

Student mobility poses another threat to achievement. Urban families living in poverty move frequently, and children from poor families are especially likely to attend schools with large numbers of new students arriving during the school year. Children attending elementary schools with considerable student mobility learn less math than do children in schools with less student turnover. Moreover, these negative effects apply to students who themselves are residentially stable as well as to those who

are not. The explanation is straightforward: integrating new students into a class reduces the time available for instruction.

Difficulty in creating strong, stable teaching faculties also contributes to the weak performance of students in high-poverty schools. Schools serving high concentrations of poor, nonwhite, and low-achieving students find it difficult to attract and retain skilled teachers. Moreover, teacher turnover reduces the payoff to investments in improving the quality and consistency of instruction.

All told, over the past 40 years, economic changes have both raised the bar for the quality of education needed to get ahead in America and made it much more difficult for schools serving low-income children to rise above it. Efforts to improve public schools often center on “silver bullets”—more money, more accountability, more choice, more charter schools. None of these initiatives has resulted in consistently better schooling for low-income children. The reason is that none focuses directly on changing students’ daily experiences in school.

It is easy to dwell on the characteristics of American education that make constructive change difficult. However, there are also strengths to build on. Of particular importance are educational interventions, conducted at considerable scale, that have been shown in rigorous evaluations to improve the skills of low-income children. Our recent book *Restoring Opportunity* features three such initiatives—the Boston pre-K program, the campuses of the University of Chicago charter school, and New York City’s small schools of choice. These innovative and quite durable programs, all of which change children’s daily school experiences, show that it is possible to improve the education of low-income children.

Although these programs are exceptional, they highlight what it takes to improve the education of low-income children on a wider scale. All take advantage of advances in research-based knowledge such as effective strategies for developing literacy skills. All provide rich opportunities for teachers and school leaders to improve their skills. All incorporate sensible systems of accountability, including teachers’ obligation to open their classrooms to the scrutiny of colleagues and school leaders. Finally, all incorporate high academic standards such as those embedded in the Common Core State Standards.

Providing high-quality education on a consistent, long-term basis to low-income children requires institutions that consistently offer supports of the same high quality as those afforded to the schools participating in the effective programs described in our book. The United States has yet to develop a set of institutions that do this effectively. In a promising recent trend, however, a growing number of innovative organizations are supporting public schools. Some, such as the New York Leadership Academy and New Leaders for New Schools, prepare principals to create schools

that are effective learning communities for both teachers and students. Others, such as Teach for America and the Boston Teacher Residency program, recruit academically talented college graduates and support their work in high-poverty schools. Still others, such as New Visions for Public Schools, the Urban Assembly, and many charter management organizations, recruit leadership teams to start new schools and provide them with ongoing support. Finally, there are the comprehensive school reform design organizations such as Success for All and America's Choice that offer detailed guidance and tools to large numbers of high-poverty schools. The challenge is to devise organizational structures that give high-poverty schools the resources, knowledge, and freedom to choose the collection of supports they need to build and sustain a high-quality, coherent educational program.

Over the past 20 years, a consensus has grown that American schools should be judged by their effectiveness in educating students. However, accountability can take different forms. In contrast to accountability systems in some states that create incentives for skilled teachers to avoid working in high-poverty schools, the greatest accountability pressure for teachers in the successful schools we highlight is to embrace the abundant opportunities to improve their teaching and to work as part of a team committed to improving the skills and life chances of every student. To improve the education of low-income children, it is essential to devise accountability systems that achieve these objectives for all schools, especially those serving high concentrations of low-income children.

In the case of the Boston pre-K program we feature in *Restoring Opportunity*, accountability came in the form of the expectation that each school would obtain and retain accreditation from the National Association for the Education of Young Children (NAEYC). For the University of Chicago charter schools and the small schools of choice in New York City we describe, accountability included the requirement that their students score well on state-mandated examinations.

But in these effective schools, teachers and school leaders experienced a more immediate and more important type of accountability: a responsibility to their colleagues for educating every student. For Boston pre-K teachers, it included taking advantage of the instructional coaching provided by the system. For the charter school teachers, it included working together to make implementation of the sophisticated literacy curriculum more consistent. For the ninth-grade teachers in New York City, it meant embracing their shared responsibility to develop the skills of all incoming students, including those reading far below grade-level. In all of these schools, teachers had the resources and supports to respond constructively to the accountability pressures.

What can schools accomplish? The answer depends on our coun-

try's commitment to creating systems that provide to all schools the combination of consistently strong supports and sensible accountability experienced by the schools we highlight. In some settings, this can be done by reallocating existing funds to serve the real needs of schools. In other settings, especially where there is a high concentration of low-income students, it will take more resources to provide all students with the high-quality educational experiences needed to improve their life chances. Creating a constituency for better schooling for low-income children requires more widespread understanding of how changes in the U.S. economy have increased the immense challenges that high-poverty schools face and an awareness that meeting these challenge is critical to the nation's future.

Is U.S. Language Education Policy Outdated?

Patricia Gándara^{1,2}

Although the United States is a self-proclaimed “nation of immigrants,” it has historically not been very welcoming of these newcomers, and especially not of their languages (Macías, 2014). As long as non-English language communities did not grow too large or become too visible, they could be accommodated into the nation. However at every point when immigration has surged and increasing numbers of people speak a language other than English, there has been a concomitant surge in English-only legislation. Today we live with language education policies at the state and federal levels that were forged in a period of very high immigration that began in the 1970s and continued until the Great Recession of 2007–2008. This paper questions whether those policies make sense in the 21st century.

A Brief History of Language Education Policy in the United States

By the end of the 19th century, German bilingual schools flourished, especially in Midwestern states, and a dozen states had passed pro-bilingual laws (Kloss, 1998). However the first decades of the 20th century saw the (proportionately) greatest immigration to the United States in its

¹ Patricia Gándara is Research Professor of Education at the University of California, Los Angeles. She was elected to the National Academy of Education in 2015.

² Some portions of the text of this essay are drawn from a forthcoming article by the author, “Charting the Relationship of English Learners and the ESEA: One Step Forward, Two Steps Back” for Russell Sage Foundation.

history, and extreme “Americanization” policies were introduced in many states. For example, in 1918 Texas passed a stringent English-only law that made it a criminal offense for teachers, principals, superintendents, and other school personnel to teach in a language other than English. In 1920, a teacher by the name of Robert Meyer was tried and convicted of teaching reading in German to a 10-year-old boy in a parochial school in Nebraska at the request of a child’s parents. Meyer was found to violate Nebraska’s prohibition of teaching any language other than English until after the 8th grade. This case eventually ended up in the Supreme Court, which found that the conviction of Meyer violated the U.S. Constitution, and Justice McReynolds added, “proficiency in a foreign language ... is not injurious to the health, morals or understanding of the ordinary child” (*Meyer v. Nebraska*, 1923). Nonetheless, by 1923, 34 states required English as the medium of instruction in the schools (Leibowitz, 1971). Teddy Roosevelt’s famous statement, “We have room for but one language in this country, and that is the English language, for we intend to see that the crucible turns our people out as Americans, of American nationality, and not as dwellers in a polyglot boarding house,” was written in 1919 (Roosevelt, 1929, p. 554), at the height of the “Americanization” movement in the United States but continues to resonate with many Americans today.

A period of benign neglect of students whose primary language was not English ensued after the collapse of the U.S. economy in the 1930s and the virtual end of immigration. However, immigration began to increase once again after the passage of the Immigration and Nationality Act of 1965, which removed country quotas, opened the doors to Asian immigration (which is now the highest for any region), and substituted the old quotas with a policy of family reunification that changed the face of the nation and increased linguistic diversity. Moreover, the passage in 1968 of the Bilingual Education Act (Title VII of the Elementary and Secondary Education Act) signaled a new interest in students who were not proficient in English and once again opened the door to officially instructing students bilingually. As one might predict based on past experience, English-only movements again began to take hold. The 1980s saw the rise of organized intolerance for any language other than English in the public sphere. “U.S. English,” a group formed to lobby against bilingual education, was founded in 1983 by a U.S. senator from California and an ophthalmologist from Michigan with strong ties to anti-immigration groups. Then in 1986 Californians passed a ballot measure making it an “Official English” state. By the end of the decade, nine more states had passed some form of official English law (King, 1997). This was followed by sometimes virulent anti-immigrant legislation in the mid-1990s and, in 1998, the near banning of bilingual instruction in California. Arizona and Massachusetts

soon followed suit with very similar laws. The anti-bilingual education movement was stopped in Colorado where voters turned down a similar measure in 2002. Thus, the period between 1992 and 2002 saw a steep increase in English-only instruction nationally—from 34 percent of English Learner (EL) students receiving an English-only program model in 1992 to 48 percent in 2002. That increase in English-only instruction has continued to the present, although current numbers are not available because the federal government has not commissioned any study of this issue since the disappearance of the Bilingual Education Act with the passage of No Child Left Behind in 2001. Today, as the number of EL students continues to grow, they are less likely to receive any primary language instruction than at any time over the past several decades, and the nation is less likely to produce biliterate individuals to join the global economy.

The Demographic Imperative

The United States is a radically different nation than it was just four decades ago when English-only legislation was beginning to sweep the country again. Today, one in four persons speaks a language other than English at home (Ryan, 2013), and the United States operates in a highly interdependent world economy. There are no more economic islands. In the 2012–2013 school year, nearly 5 million students across the United States were designated as ELs, comprising almost 10 percent of the total school-age population (Zong & Batalova, 2015). However, many students who today do not carry the label of EL were once ELs and are probably still on a continuum of learning academic English. Moreover, most of these students who were once ELs typically go home each day to an environment in which English is rarely heard. Because there is no national test of English proficiency or even agreement as to what constitutes “proficiency” in English for academic purposes, any count of the number of ELs is in reality a best estimate. Considering all the students who are coming from homes where English is not the primary language, the total percentage of students nationally who are still learning to learn in English is probably more like 20 percent. While these children are often referred to as “immigrant children,” actually relatively few ELs are born outside the country. In 2013, 88 percent of children of immigrant parents were native-born citizens (Zong & Batalova, 2015).

Language Education Policies

In a seminal article describing the “orientations” or philosophies of policymakers regarding language policies, Richard Ruiz (1984) contends that there are three basic orientations: (1) language as a problem, some-

thing that needs to be fixed; (2) language as a right, something that must be legally protected; and (3) language as a resource, something that is an asset. The United States has largely adopted a language-as-a-problem orientation and, by doing so, practices result that focus neither on social justice (a right) nor on asset development (a resource) and overlook and under-estimate the assets that immigrant students and ELs possess.

Framing English Learners

Immigrant students and those students who are born in the United States but arrive at school with a primary language other than English are usually defined by what they are lacking: English language skills. This framing of the students has resulted in them being viewed as deficient, remedial, or lacking in fundamental skills that are critical for normal academic achievement. As such, most programs that serve these students are geared to remediate their deficiency, and only once this is accomplished are they deemed ready to join the mainstream and have full access to a regular curriculum (Callahan, 2005; Callahan & Gándara, 2004). The evidence suggests that this policy has not served these students well. Not only are they denied development of their native language, but also there has been no real narrowing of the gap in academic skills in English with their English-speaking peers over the past couple decades.

Nationally, since 1996 (the first year for which the National Assessment of Education Progress [NAEP] shows the gap trend lines for ELs), there has been no closing of the gap in either 4th- or 8th-grade math, and in fact, the gaps have begun to widen between ELs and all others. For example, in 2003 the gap between English learners and English speakers who scored proficient in 8th-grade mathematics was 20 points; in 2013 the gap had grown to 24 points. Eighth-grade reading proficiency showed a similar widening of the gap (3 points) over the same period. At least from the perspective of math and reading score gaps, educational achievement has not improved nationally for ELs, and across the grades these students remain significantly behind their native English-speaking peers. Another important indicator of academic success is high school graduation. Current estimates are that students who are categorized as EL in secondary school are twice as likely to drop out of high school as all other students (Callahan, 2013).

Why English-Only Instruction?

Given that existing language education policies do not appear to be particularly successful, why is there such adherence to English-only instruction, and what are the arguments against bilingual instruction?

During the 1980s—the same time period in which anti-immigrant sentiments were growing intense across the nation—bilingual education was pronounced a failure. Two influential studies had failed to find superior outcomes for EL students enrolled in bilingual compared to EL students in English-only programs. The first of those studies, conducted by the American Institutes for Research in 1977–1978 (Danoff, 1978) compared students in 38 Title VII programs with similar students in English-as-a-Second-Language (presumably English only) classrooms over a period of several months. The study was criticized by many researchers for including programs in the two groups solely on the basis of program labels without examining the actual educational treatment provided or controlling for differences in the students assigned to the programs. These, and other methodological problems, called the findings into question. Nonetheless, the study was held up by anti-bilingual partisans as evidence of bilingual education's failure.

The second highly influential study was a “meta-analysis” conducted by Keith Baker and Adriana de Kanter (1981) of the U.S. Department of Education that reviewed 28 studies that met their definition of methodological rigor to be included in a *qualitative* analysis of the programs—“yes” the evaluation found positive effects for bilingual instruction or “no” it did not. They did not attempt to quantify the degree of effectiveness, and they had no first-hand knowledge of the “treatments.” Of course, the findings of any meta-analysis turn on the programs selected for study, and this study was widely criticized for the programs selected in addition to the crude methods used. Notwithstanding these critiques, the study concluded that, “Evidence does not support the necessity of teaching non-language subjects in the child's native tongue” (p. 1).

What was left unsaid about these studies and others during the period was that neither English-only nor bilingual programs was found to be superior to the other, and therefore a conclusion that English only had somehow “won the contest” was unwarranted. However, they provided the grounds for opponents of bilingual education to say they had research on their side and “bilingual education was a failure.”

If bilingual education could not be shown to be superior to English-only methods, then the question of cost became another argument used against primary language programs. In fact, even the language of the anti-bilingual initiative in California (and later in Arizona and Massachusetts) invoked cost as an important reason to curtail these programs:

The public schools of California currently do a poor job of educating immigrant children, *wasting financial resources on costly* [emphasis added] experimental language programs whose failure over the past two decades is demonstrated by the current high drop-out rates and low

English literacy levels of many immigrant children. (Preamble to Proposition 227 of 1998)

However, while the research on the costs of bilingual education versus other language assistance programs is notoriously subjective and variable (see Rumberger & Gándara, 2015), two studies reported in the literature are notable for relying on actual district-level data. These studies of the cost of bilingual education compared to other program alternatives found that because the bilingual teacher is able to provide all of the instruction for his or her students, while other models generally rely on additional personnel (e.g., resource teachers, special bilingual aides) to assist with instruction, bilingual education can be more cost-effective (Huffman & Samulon, 1981; Parrish, 1994). So, it turns out that the cost issue is a red herring. More likely, the reactions against bilingual education were related to the same recurring theme in American history: when immigration increases to levels that the general population begins noticing it, there are reactions against it, including calls for immigrants to “speak English.”

A New Generation of Research on Bilingualism and Its Benefits

Both the knowledge base regarding language acquisition and methodological techniques for studying it have developed substantially over the past several decades. Newer studies have addressed many of the limitations of earlier research. In a best evidence meta-analytic study, Robert Slavin and Alan Cheung (2005) of Johns Hopkins University found that among the 17 studies that met their strict methodological criteria for inclusion, 13 favored bilingual programs (all Spanish-English) and 4 found no differences. This study, in contrast to the earlier Baker and de Kanter study incorporated quantitative methods to determine the actual effect sizes of the treatments. The effect size for the averaged score differences was between .33 and .45, indicating a “medium positive effect.” Across both the Slavin and Cheung study and four other rigorous meta-analyses, August, Goldenberg, and Rueda (2010) found “differences in favor of native-language instruction, with effect sizes ranges from small to moderate” (p. 143). They also noted that the better the technical quality of the studies, the larger were the effect sizes. In a synthesis of the most rigorous research on reading instructional approaches for ELs, Claude Goldenberg (2008) also concluded that teaching students to read in their first language promotes higher levels of reading achievement in English, a finding that is often thought to be counterintuitive, although it is well supported by theory as well as by data. The theories underlying this finding are that of *transfer* (knowledge acquired in one language is transferred to additional languages as they are learned) and *comprehensible input* (indi-

viduals learn more efficiently when they can understand at least part of what is being communicated) (Cummins, 1981; Krashen, 1987).

Most evaluation research on bilingual education has focused narrowly on short-term outcomes for reading and sometimes math in English. Very little attention has been paid to longer-term effects or to other potential outcomes. In fact, many of the studies that have found no difference or less positive effects for bilingual instruction have been based on very short-term analyses. Genesee, Lindholm-Leary, Saunders, & Christian (2006), reporting on a synthesis of research on ELs, note:

Evaluations conducted in the early years of a program (Grades K-3) typically reveal that students in bilingual education scored below grade level ... [but] Almost all evaluations of students at the end of elementary school and in middle and high school show that the educational outcomes of bilingually educated students, especially those in late-exit and two-way programs, were at least comparable to and usually higher than their comparison peers. (p. 201)

A recent study that followed thousands of students in one large school district who were assigned to transitional bilingual (i.e., early exit), dual language bilingual (longer term, incorporating English speakers and English learners), and English-only programs beginning in kindergarten and following them into high school found that the EL students who had remained in bilingual instruction, and especially two-way dual language programs, out-performed the students in English-only instruction on all measures. They ultimately reclassified to English proficient at higher rates and scored higher on English Language Arts and on measures of English proficiency (Umansky & Reardon, 2015). Because all observable characteristics that could affect student outcomes, such as socioeconomic status, race and ethnicity, and prior test scores in English had been controlled, the study can be considered particularly robust.

With respect to outcomes other than test scores or English proficiency, there is now a large and growing body of research on a host of other outcomes, including both cognitive and noncognitive. Bialystok and her colleagues (Bialystok, 2001; Bialystok & Craik, 2010; Bialystok & Majunder, 1998) found in a series of innovative studies that bilingually educated students tend to have greater cognitive flexibility, working memory, and executive functioning (e.g., concentration) than monolinguals. Portes and Hao (1982) found that bilingual students from immigrant families have more cohesive family relations and fewer behavior problems in school. They attribute this, as have others, to greater communication and parental authority fostered by parents and children communicating in the same language. Genesee and Gándara (1999), in a review of the impact of bilingual education on intergroup relations found that bilingual programs that

mix English speakers with EL students produce students who are more open to other cultures and more likely to create friendships with students with different cultural backgrounds than their own. These are characteristics that employers say they value in their workers (Forbes Insight, 2012).

Bilingualism and the U.S. Labor Market

Perhaps somewhat surprisingly, the accumulating evidence has done relatively little to change policy toward language education at either the federal or state level. Given this fact, the Civil Rights Project at the University of California, Los Angeles, undertook to determine whether bilingually educated individuals are compensated for this skill in the U.S. labor market. Because money is a powerful incentive in this society, perhaps answering this question positively would attract the attention of education policymakers.

What we found initially was not encouraging. A number of studies have been undertaken on the impact of bilingualism on labor market earnings by economists, and there is a consensus among these studies that bilingualism not only does not pay, but also it may exact an earnings penalty. Bilinguals tend to earn less than monolinguals in the same or similar positions (Callahan & Gándara, 2014). After commissioning a series of studies to test these findings we continued to find the same conclusion: being a bilingual in the U.S. labor market did not carry a wage premium, and in fact, often carried a wage penalty (Alarcón, DiPaolo, Heyman, & Morales, 2014; Robinson-Cimpian, 2014). While initially puzzling, this finding could be explained by considering who is bilingual in the United States? Immigrants. And immigrants generally earn less than native-born citizens (Anderson, 2015). But clearly more research was needed.

We convinced several researchers to investigate questions of potential economic advantage for bilinguals using data sets that could yield more specific information about the bilingual's level of literacy in both languages. Using the National Educational Longitudinal Study (NELS:88) data set, Agirdag (2014) controlled for a host of variables associated with achievement outcomes, including socioeconomic status, prior test scores, and several noncognitive variables and compared earnings for the balanced bilinguals (those who were literate in both languages) with English-dominant individuals at about age 26. He found that not only did the balanced bilinguals earn more in the labor market overall but also losing one's primary language carried a cost of about \$2,000 to \$3,000 annually in 2012 dollars.

Lucrecia Santibañez and Maria Estela Zárate (2014), analyzed longitudinal data from the Educational Longitudinal Study (ELS) of the U.S. Department of Education, and found that students from bilingual homes

who maintain their bilingualism into high school are more likely to go to college than those who lose the home language, and that balanced bilingual Latinos are more likely to go to 4-year colleges. Both of these findings are, of course, of great interest because Latinos are the least likely of all major subgroups to complete a college degree, and completing the degree is closely tied to a substantial increase in earnings (Baum, Ma, & Payea, 2013).

Rubén Rumbaut (2014) conducted a series of analyses of two merged longitudinal data sets with more than 6,000 youth and young adults from the Southern California region. He found that those young adults from immigrant backgrounds who maintained balanced bilingual skills are less likely to drop out of high school and more likely to secure higher level positions in the workforce and earn more at those jobs than monolinguals or weaker bilinguals.

Conclusions

While English is undoubtedly the most valued world language, the person who speaks English *in addition to* other languages will always have an advantage in a globalized world. The United States, with its wealth of cultures and languages, is enormously advantaged in this regard. However, as the research has also shown, this is a small window of opportunity. Immigrants and their children are losing the first language more rapidly today than in the past (National Research Council, 2010; Rumbaut, 2009).

There is another lesson from this body of new research: the emphasis on *transitional* bilingual instruction (transferring to all English at the first opportunity) is not where the economic payoff occurs. The economic payoff for both individuals and society, it turns out, comes from maintaining and developing home languages. We need to reframe the ELs and immigrant students in our society as a source of enormous advantage for the nation, and to educate them accordingly.

References

- Agirdag, O. (2014). The literal cost of language assimilation for the children of immigration: The effects of bilingualism on labor market outcomes. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage. Language, literacy, and the U.S. labor market* (pp. 160–181). Bristol, UK: Multilingual Matters.
- Alarcón, A., DiPaolo, A., Heyman, J., & Morales, M. C. (2014). Returns to Spanish-English bilingualism in the new information economy: The health and criminal justice sectors in the U.S. borderlands with Mexico. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage. Language, literacy, and the U.S. labor market* (pp. 110–137). Bristol, UK: Multilingual Matters.

- Anderson, K. (2015). "Can immigrants ever earn as much as native workers?" IZA World of Labor. doi: 10.15185/izawol.159. Retrieved from <http://wol.iza.org/articles/can-immigrants-ever-earn-as-much-as-native-workers.one-pager.pdf>.
- August, D., Goldenberg, C., & Rueda, R. (2010). Restrictive state language policies: Are they scientifically based? In P. Gándara & M. Hopkins (Eds.), *Forbidden language: English learners and restrictive language policies*. New York: Teachers College Press.
- Baker, K., & de Kanter, A. (1981). *The effectiveness of bilingual education. A review*. Final Draft Report. Washington, DC: U.S. Department of Education. Retrieved from <http://files.eric.ed.gov/fulltext/ED215010.pdf>.
- Baum, S., Ma, J., & Payea, K. (2013). *Education pays. The benefits of higher education for individuals and society*. New York: The College Board.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. New York: Cambridge University Press.
- Bialystok, E., & Craik, F. (2010). Cognitive and linguistic processing in the bilingual mind, *Current Directions in Psychological Science*, 19, 19–23.
- Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving, *Applied Psycholinguistics*, 19, 69–85.
- Callahan, R. M. (2005). Tracking and high school English learners: Limiting opportunity to learn. *American Educational Research Journal*, 42, 305–328.
- Callahan, R. (2013). The English learner dropout dilemma: Multiple risks and multiple resources. Santa Barbara, CA: California Drop out Research Project. Retrieved from http://www.cdpr.ucsb.edu/pubs_reports.htm.
- Callahan, R. M., & Gándara, P. (2004). On nobody's agenda: English learners and postsecondary education. In M. Sadowski (Ed.), *Immigrant and English-language learners: Strategies for success* (pp. 107–127). Cambridge, MA: Harvard Education Press.
- Cummins, J. (1981). The role of primary language development in promoting educational success for language minority students. In *Schooling and language minority students: A theoretical framework*. Los Angeles, CA: California State University Evaluation, Dissemination, and Assessment Center.
- Danoff, M. (1978). *Evaluation of the impact of ESEA Title VII Spanish/English bilingual education programs*. Washington, DC: American Institute for Research. ERIC document no. ED154634.
- Forbes Insights. (2011). Reducing the impact of language barriers. Retrieved from http://www.forbes.com/forbesinsights/language_study_reg/index.html www.forbes.com/forbesinsights.
- Gándara, P., Alvarado, E., Driscoll, A., & Orfield, G. (2011). *Building pathways to transfer: Community colleges that break the chain of failure for students of color. A report to the Ford Foundation*. Los Angeles, CA: Civil Rights Project, UCLA.
- Genesee, F., and Gándara, P. (1999). Bilingual education programs: A cross-national perspective, *Journal of Social Issues*, 55, 665–685.
- Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2006). *Educating English learners: A synthesis of research evidence*. Cambridge, MA: Cambridge University Press.
- Goldenberg, C. (2008, Summer). Teaching English language learners. *American Educator*, 8–23, 42–43.
- Huffman, P. C., & Samulon, M. (1981). *Case studies of delivery and cost of bilingual education: A Rand note*. Santa Monica, CA: The RAND Corporation.
- Jenkins, V. (2013, December 29). Cumberland County schools dual immersion programs score high. *FayObserver.com*. Retrieved from http://www.fayobserver.com/news/local/article_6297cba0-a357-5c77-916a-ece9b2eb109e.html.
- King, R. (1997, April). Should English only be the law? *The Atlantic Monthly*.
- Kloss, H. (1998). *The American bilingual tradition*. Rowley, MA: Newbury House Publishers.

- Krashen, S. (1987). *Principles and practice in second language acquisition*. New York: Prentice Hall.
- Leibowitz, A. (1971, March). *Educational policy and political acceptance: The imposition of English as the language of instruction in American Schools*. ERIC document no. ED 047321.
- Macias, R. (2014). Benefits of bilingualism: In the eye of the beholder. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage. Language, literacy, and the U.S. labor market* (pp. 16–44). Bristol, UK: Multilingual Matters.
- Moore, S., Fee, M., Ee, J., Wiley, T., & Arias, B. (2014). Exploring bilingualism, literacy, employability and income levels among Latinos in the United States. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage. Language, literacy, and the U.S. labor market* (pp. 16–44). Bristol, UK: Multilingual Matters.
- National Research Council. (2006). *Multiple origins, uncertain destinies: Hispanics and the American future*. M. Tienda and F. Mitchell (Eds.), Panel on Hispanics in the United States, Committee on Population, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- Parrish, T. (1994). A cost analysis of alternative instructional models for Limited English Proficient Students in California. *Journal of Education Finance*, 19, 256–278.
- Portes, A., & Hao, L. (2002). The price of uniformity: Language, family, and personality adjustment in the immigrant second generation, *Ethnic and Racial Studies*, 25, 889–912.
- Proposition 227. California ballot initiative. June 2, 1998. Retrieved from [http://ballotpedia.org/California_Proposition_227_the_%22English_in_Public_Schools%22_Initiative_\(1998\)](http://ballotpedia.org/California_Proposition_227_the_%22English_in_Public_Schools%22_Initiative_(1998)).
- Robinson-Cimpian, J. (2014). Labor market differences between bilingual and monolingual Hispanics. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage: Language, literacy and the U.S. labor market* (pp. 79–109). Bristol, UK: Multilingual Matters.
- Roosevelt, T. (1926). *Works* (Memorial edition), Vol. XXIV. New York: Charles Scribner's Sons.
- Ruiz, R. (1984). Orientations in language planning, *NABE Journal*, 8, 15–34.
- Rumbaut, R. (2014). English plus: Exploring the socio-economic benefits of bilingualism in Southern California. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage: Language, literacy and the U.S. labor market* (pp. 182–210). Bristol, UK: Multilingual Matters.
- Rumberger, R., & Gándara, P. (2015). Resource needs for educating linguistic minority students. In H. Ladd & M. Goertz (Eds.), *Handbook of research in education finance and policy* (pp. 585–606). New York and London: Routledge.
- Ryan, C. (2013). *Language use in the United States: American Community Survey reports* (Vol. ACS-22). Washington, DC: U.S. Census Bureau.
- Santibañez, L., & Zárte, M. E. (2014). Bilinguals in the United States and college enrollment. In R. Callahan & P. Gándara (Eds.), *The bilingual advantage: Language, literacy and the U.S. labor market* (pp. 211–233). Bristol, UK: Multilingual Matters.
- Slavin, R., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English Language Learners. *Review of Educational Research*, 75, 247–284.
- Umansky, I., & Reardon S. (2014). Reclassification patterns among Latino English learner students in bilingual, dual immersion, and English immersion classrooms. *American Educational Research Journal*, 51, 871–912.
- Zong, J., & Batalova, J. (2015). *Frequently requested statistics on immigrants and immigration in the United States*. Washington, DC: Migration Policy Institute. Retrieved from <http://www.migrationpolicy.org/article/frequently-requested-statistics-immigrants-and-immigration-united-states>.

Revisiting the High Hopes and Broken Promises of Public Education: Still an Uncertain Future

*Sonia Nieto*¹

When National Academy of Engineering (NAEd) members were asked to consider writing an essay for the 50th anniversary of the organization, I accepted the invitation because I believe that these occasions give us an opportunity to assess progress and think about what still needs to be done. I have been fortunate to have had two such opportunities. In 2005, I was invited by the editors of the *Harvard Educational Review* (HER) to write an article for the 75th anniversary issue of the journal. For that anniversary, I decided to explore how close our nation was to fulfilling its promise of equal education for all children (Nieto, 2005). The resulting article, “Public Education in the Twentieth Century and Beyond: High Hopes, Broken Promises, and an Uncertain Future,” chronicled the advances, unfulfilled potential, and hopes for a better future in public education.

Ten years later, in this essay I revisit that article to explore how far we have come since 2005 by addressing the current state of equality and equity in public education. As the subtitle of this NAEd volume proclaims, the past is indeed prologue.

¹ Sonia Nieto is Professor Emerita at University of Massachusetts, Amherst. She was elected to the National Academy of Education in 2015.

High Hopes and Broken Promises

In my 2005 HER article, I recalled the high hopes that many in our nation have had for public education since the 19th century. Notably, Horace Mann, in his 12th annual report to the Massachusetts State Board of Education, confidently declared, "Education then, beyond all other devices of human origin, is a great equalizer of the condition of men" (Mann, 1868, p. 669). In that article, I also cited John Dewey, whose ideas about education at the dawn of the 20th century cemented the connection between education and democracy (see especially Dewey, 1916). Dewey viewed schools as not only serving the public good, but also as an apprenticeship for civic life. The common school, Dewey hoped, would make visible the nation's commitment to "a more perfect union" where the responsibilities and rights of citizenship would be available to all. While access to public education since Dewey's time has increased greatly—females are now assured a seat in U.S. classrooms, as are children of color and immigrants—but even though educational outcomes have improved for large segments of the population, public education today is in danger of losing its historic link to democracy and the public good.

Even in the few years since my HER article was published, we have witnessed a dramatic dismantling of public education primarily because of two factors—segregation and privatization. The growing segregation, or resegregation, of U.S. schools, is evident in proportions not seen since before the historic 1954 *Brown v. Board of Education* decision (Orfield, Kucsera, & Siegel-Hawley, 2012). While some politicians and policymakers wax poetic about the significance of *Brown*, little has actually changed in terms of providing equal access to children of color, particularly those living in poverty. In fact, the case has been persuasively made that conditions for African American children have actually worsened since the *Brown* decision (Anderson, 1988; Ladson-Billings, 2006; Walker, 2000). In addition, in terms of ethnicity and social class, Latin@s are now the most segregated group of children in U.S. schools (Frankenberg & Orfield, 2007).

The other leg of the so-called reform movement, privatization, has come in the form of charter schools and for-profit education. In an extensive review of the research on charter schools, and in spite of the hype about their effectiveness and the bipartisan support they enjoy, David Berliner and Gene Glass found that charter schools are no more effective than non-charter schools and that, in some cases, they are decidedly worse (Berliner & Glass, 2014). The real crisis in education, Berliner and Glass contend, is the brutal inequality in our society, which has grave implications for children living in poverty, not only for their education but also for every aspect of their lives. In a recent article on the dramatic

increase in charter schools, Thandeka Chapman and Jamel Donnor (2015) use a CRT (Critical Race Theory) lens to assess the impact of charter schools. As a result of their investigation, Chapman and Donnor challenge the notion of marketplace theory as a viable reform strategy to equalize education. Instead, they maintain that the substantial financial profits associated with charter schools result in policymakers turning a blind eye to the significant research on their negative outcomes.

Privatization brings up another dilemma: thinking of education as simply another commodity threatens the very basis of civic life in a democratic society. For instance, vouchers, now more plentiful than at any time in U.S. education, are changing the meaning and nature of public education. Michael Apple (2006) terms the neoliberal and neoconservative policies of the past three decades "conservative modernization." He explains, "Conservative modernization has radically reshaped the common sense of society. It has worked in every sphere—the economic, the political, and the cultural—to alter the basic categories we use to evaluate our institutions and our public and private lives" (Apple, 2006, p. 226). In fact, these policies pose a direct challenge to the democratic purposes of education. In a book on the enormous spread of vouchers in Minneapolis, some even for religious schools, Barbara Miner bemoaned the fact that "In the current debates on vouchers, there is strikingly little discussion between democratic values, the common good, and public education" (Miner, 2013, p. 174).

It is not only policies, programs, and practices that have failed our young people during the past decade, but also a dismissive discourse about public education that privileges functionalism over creativity and rigid accountability over the joy of learning. Nowadays we are more likely to hear words that bring to mind punishment and control rather than those that refer to the excitement of learning. A poignant example comes from an essay written by Greg Michie about his return to teaching. After being in higher education for more than a decade, he wrote about his return to the classroom. Greg thought he was prepared for the changes that had taken place in public education. Though he had been a teacher educator for a dozen years, he had spent many hours in teachers' classrooms in those years. But even a veteran classroom connoisseur such as Greg was not ready for what he found. He wrote:

I knew I'd probably hear far less in the coming year about democratic education or social justice than I would about current buzzwords like "text complexity" "accountable talk," and "close reading." But even though I understood all this going in, experiencing it from the ground level perspective of a teacher was still jarring.... This isn't what I came back to do, I thought to myself. I couldn't remember the word "data" even being mentioned during my previous tenure as a teacher. Now, it

was the centerpiece of discussion, the tail that wagged the dog. (Michie, 2015, p. 114)

Desegregation, Bilingual Education, and Multicultural Education

The “high hopes” to which I referred in my previous article focused specifically on three movements in education—desegregation, bilingual education, and multicultural education—all of which were a direct result of the 1960s civil rights movement. Many activists viewed these movements as a remedy to what Jonathan Kozol had aptly termed “savage inequalities” in schooling (Kozol, 1991).

Families and communities whose children had been poorly served by public education, particularly children of color, language minority children, and immigrants, seized upon these movements to offset the failures of public education. To provide a context for the discussion, I explored the rise and fall of these movements. All three had their heyday in the 1960s and 1970s. It is no coincidence that it was as a result of Great Society programs during those years that the nation experienced the greatest progress in equalizing education, nor is it a surprise that in the 1980s these successes began to be reversed with the conservative and neoconservative turn in politics that also greatly influenced education policy.

The fact that race, ethnicity, social class, and language were central to these three movements is also not coincidental. For one, it has always been the marginalized in our society rather than the privileged who have advocated for change, for “a piece of the pie,” and for more equitable educational outcomes that might lead to better lives. This explains the struggle for literacy among African Americans, the demand for integrated schools for Mexican children who were forced to attend so-called Mexican schools, the repudiation by American Indians of the brutal policy of forcing their children into boarding schools, the press for schools that would serve young people with special needs, and other struggles related to education. People whose children already enjoy the privilege of an excellent and high-quality education are not always eager to share this privilege with less privileged communities.

In addition, the racial reality of our society had changed substantially in the decades since the 1950s when white children were the great majority in U.S. schools. The annual 2015 *Condition of Education* found that between just 2002 and 2012, the number of white children in U.S. public schools decreased markedly from 28.6 million to 25.4 million, and their percentage decreased from 59 to just 51 percent. The number of African American students has remained about the same, decreasing slightly from 17 to 16 percent, while Asian and Pacific Island students have increased from 4 to 5 percent. By far the greatest increase has been among Latin@

students, from 8.6 to 12.1 million, or from 18 to 24 percent (U.S. Department of Education, 2015).

The hopes spurred by the civil rights movement, the growing diversity of the student population, and the ongoing dissatisfaction with the failure of schools to give children of color, language minority students, and immigrants a chance at an equal education all had a hand in creating and sustaining the movements for desegregation, bilingual education, and multicultural education. Their victories, however, were short-lived, and today this is more apparent than ever. For example, desegregation is hardly mentioned anymore as a solution to unequal education (Frankenberg & Orfield, 2007). As a result, today there are few examples of truly integrated schools, and these are primarily magnet schools. Also, increased residential segregation has resulted in schools becoming more racially and socioeconomically segregated than ever (Orfield, 2009). Yet, according to Orfield, desegregation of public schools offers the single most powerful way to prepare the coming generation for a multiracial society, one that will have no racial majority group (Orfield, 2009).

Bilingual education, although always controversial, enjoyed some measure of support during the 1970s. Today, save for the few states where bilingual education is thriving, the words “bilingual” and “education” are rarely heard together. They have instead been replaced by “Education for English Language Learners” in city and state departments of education, as well as in the federal government. It has become politically correct nowadays to disavow bilingual education in favor of ESL (English as a Second Language) or education for English Language Learners (ELLs), as if the only task for these children is to learn English. While English monolingualism is lauded as an appropriate outcome for immigrant students, becoming bilingual has come to be seen as a deficit rather than as the asset it is (Garcia & Klieffen, 2010). Several studies and books have examined the decline of bilingual education as a result of voter initiatives in a number of states, with a concomitant decline in educational outcomes for immigrant and other students for whom English is an additional language (Gándara & Contreras, 2009; Gutierrez, Baquedano-Lopez, & Asato, 2001; Tung et al., 2009).

Multicultural education, a movement that began in earnest in the mid-1970s, has made an enormous contribution to both K–12 education and teacher preparation (Banks & Banks, 2004). It has also seen its share of controversy (Sleeter, 1995). The scholarship and interest in the field grew tremendously from the 1970s to the present, with almost all colleges of education offering or requiring their preservice students to take at least one course on the topic. Nevertheless, given the so-called education reform movement of the past three decades, it has been difficult for schools and teachers to implement a multicultural perspective to any

great extent in most public schools. With calls for more accountability and a more rigidly standardized curriculum, the situation has worsened in the past decade. And although scholars such as Christine Sleeter have made the important point that high standards are compatible with multicultural education (Sleeter, 2005), the suffocating presence of high-stakes testing and other standardizing policies have made this a difficult sell.

An Uncertain Future

When I was a young teacher, I believed fervently in the dream that education, if not “the great equalizer,” could at least be an equalizer of sorts. After all, that was the case for me. As a first-generation, working-class Puerto Rican, I know that were it not for public education, I would not be where I am today. At the same time, I recognize that many young people, including some of my own family members and many peers, have not been as fortunate. I recognize as well that it was only because my family moved out of the “wrong” zip code when I was an adolescent that I was able to receive the excellent education that, despite the shortcomings of my previous years in poor schools, was able to compensate and prepare me for a more consequential future than would otherwise have been the case.

Today, many children continue to be failed by public education because of their race, ethnicity, immigration status, native language, social class, other difference, or simply because they reside in the “wrong” zip code. All of these are beyond the power of children to change, and yet our society insists on blaming them, their parents, and their communities for their failure. The result of our failed education policies is the use of descriptors such as “underperforming schools,” “probation,” and other punitive terms about schools attended by young people who are especially vulnerable and need the most support.

As I did in my 2005 article, I question where we have been, and I wonder about the future. What have we learned in the past decade? Has education improved significantly for those who have been least well served by public education? Is the future of our young people brighter now than it was then? I end this essay with an unfortunate conclusion: despite some modest progress, I believe the future of public education is even more uncertain today than it was a decade ago, particularly for our most marginalized and educationally oppressed children. In this sense, inequality in U.S. schools has not changed substantially. The words of Italian educator and fierce advocate for the education of immigrant students of all backgrounds, Francesco Cordasco, ring true even today, more than 40 years after he first wrote them:

In a multi-racial, ethnically variegated society, the American experience (certainly in its schools) has been an experience of cultural assault, discriminatory rejection of educational opportunity for many children, and the continuation of social and economic advantage for a white Anglo-Saxon, Protestant, middle-class patrician elite. (Cordasco, 1973, 1998, p. 4)

How our nation addresses this issue in the coming decades will say more about who we are, what we value, and the hopes we have for the future more powerfully than anything else we do as a society. The role of the National Academy of Education can be a crucial one in this endeavor.

References

- Anderson, J. D. (1988). *The education of Blacks in the south, 1860–1935*. Chapel Hill, NC: University of North Carolina Press.
- Apple, M. W. (2006). *Educating the “right” way: Markets, standards, God, and inequality* (2nd ed.). New York: Routledge.
- Banks, J. A., & Banks, C. A. M. (Eds.). (2004). *Handbook of research on multicultural education* (2nd ed.). San Francisco, CA: Jossey-Bass.
- Berliner, D. C., & Glass, G. V. (2014). *50 myths and lies that threaten America’s public schools: The real crisis in education*. New York: Teachers College Press.
- Chapman, T. K., & Donnor, J. K. (2015). Critical Race Theory and the proliferation of U.S. charter schools. *Equity & Excellence in Education*, 48(1), 137–157.
- Cordasco, F. (1998). America and the quest for equal educational opportunity: The schools and the children of the poor. *Selected documents in social policy* (reprinted from the *British Journal of Educational Studies*, 21, February, 1973, 0–63). New York: Edna Vaughn.
- Dewey, J. (1916). *Education and democracy*. New York: Free Press.
- Frankenberg, E., & Orfield, G. (2007). *Lessons in integration: Realizing the promise of racial diversity in American schools*. Charlottesville, VA: University of Virginia Press.
- Gándara, P., & Contreras, F. (2009). *The Latino education crisis: The consequences of failed social policies*. Cambridge, MA: Harvard University Press.
- Garcia, O., & Klieffen, J. (2010). *Educating emergent bilingual: Policies, programs, and practices for English language learners*. New York: Teachers College Press.
- Gutiérrez, K., Baquedano-Lopez, P., & Asato, J. (2001). English for the children: The new literacy of the old world order. *Bilingual Review Journal*, 24(1 & 2), 87–112.
- Kozol, J. (1991). *Savage inequalities: Children in America’s schools*. New York: Crown.
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3–12.
- Mann, H. (1868). Twelfth annual report to the Massachusetts State Board of Education, 1848. In M. Mann (Ed.), *Life and works of Horace Mann* (vol. 3, p. 669). Boston, MA: Walker, Fuller.
- Michie, G. (2015). Same as it never was: On my return to teaching. In S. Nieto (Ed.), *Why we teach now* (pp. 112–121). New York: Teachers College Press.
- Miner, B. J. (2013). *Lessons from the heartland: A turbulent half-century of public education in an iconic American city*. New York: The New Press.
- Nieto, S. (2005). Public education in the twentieth century and beyond: High hopes, broken promises, and an uncertain future. *Harvard Educational Review*, 75(1), 57–78.

- Orfield, G. (2009). *Reviving the goal of an integrated society: A 21st century challenge*. Los Angeles: The Civil Rights Project/Proyecto Derechos Civiles. University of California, Los Angeles.
- Orfield, G., Kucsera, J., & Siegel-Hawley, G. (2012). *E pluribus... Separation: Deepening double segregation for more students*. Los Angeles: The Civil Rights Project/Proyecto Derechos Civiles. University of California, Los Angeles.
- Sleeter, C. E. (1995). An analysis of the critiques of multicultural education. In J. A. Banks & C. A. M. Banks (Eds.), *Handbook of research on multicultural education* (pp. 81–94). New York: Macmillan.
- Sleeter, C. E. (2005). *Un-standardizing curriculum: Multicultural teaching in the standards-based classroom*. New York: Teachers College Press.
- Tung, R., Uriarte, M., Diez, V., Lavan, N., Agusti, N., Karp, F., & Tatjana, M. (2009). *English learners in Boston public schools: Enrollment, engagement and academic outcomes, AY 2003–2006 Final Report*. Boston, MA: Gastón Institute Publications.
- U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD). (2015). State nonfiscal survey of public elementary and secondary education, 2002–03 and 2012–13; and National Elementary and Secondary Enrollment Projection Model, 1972 through 2024. Retrieved from http://nces.ed.gov/programs/coe/pdf/coe_cge.pdf.
- Walker, V. S. (2000). Valued segregated schools for African American children in the South, 1935–1969: A review of common themes and characteristics. *Review of Educational Research*, 70, 253–286.

Assimilation Without a Blueprint: Children of Immigrants in New Places of Settlement

*Alejandro Portes*¹

The New Second Generation

Over the past 25 years, a series of major studies have been conducted on the adaptation of the second-generation children of immigrants born in the United States or brought from a young age from abroad. These studies have produced the current state of knowledge on this population and have given rise to a series of theories explaining its present condition and future prospects. Their overall importance is to provide a handle on how to understand and address the situation and adaptation problems of the fastest-growing segment of America's young population, ages 24 or under. Approximately one-fourth of the country's young population in this age bracket is composed of immigrant children or children of immigrants.

Since the late 1990s, a new development has caught the attention of academics and policymakers in the field of immigration. This is the large-scale displacement of settled immigrants and new arrivals from traditional areas of concentration in the West and Southwest toward new destinations in the Midwest, East, and South. Mexicans, by far the largest immigrant group in the United States, as well as Guatemalans, Salvadorans, Hondurans, and others began to make their presence felt in cities and rural areas that had previously been the preserve of the native-born, either

¹ Alejandro Portes is the Howard Harrison and Gabrielle S. Beck Professor of Sociology (Emeritus) at Princeton University and Research Professor at University of Miami. He was elected to the National Academy of Education in 2015.

white or black. States in these regions saw their foreign-born population increase by remarkable numbers. Between 1990 and 2010, that population grew fastest in Southern and Midwestern states, including North Carolina (525 percent), Tennessee (389 percent), Iowa (301 percent), Nebraska (298 percent), and others (Portes & Rumbaut, 2014).

Naturally, the initial focus of this literature was on adult immigrants because they were the most visible part of this population. However, it became increasingly evident that it also included a large component of children because many adult immigrants brought their families with them. Unaccompanied children have also been sent by relatives at home to join parents already in the United States. This phenomenon has been rapidly accelerating during recent years. The available evidence indicates that the adult immigrant population in new places of settlement is primarily composed of low-education, low-skilled workers. Their offspring, coming from these disadvantaged family backgrounds, have made their presence increasingly felt in the public schools in these areas.

In Marshalltown, Iowa, a city of about 25,000, the foreign-born student population in the local high school went from near zero to 25 percent of the student body in the course of a decade. Preliminary interviews with several of these students, mostly Mexican-born youths brought at an early age to the United States, suggested several important trends. First, they were mostly undocumented and worried about their own future after leaving high school; second, they spoke mostly English and, when questioned in Spanish, revealed a halting command of the language indicating rapid acculturation; and third, they reported being subjected routinely to racial slurs in the streets and stores of the town. Remarks such as “go back to Mexico, you don’t belong here” were frequently addressed to them.²

Although these are initial impressions, it is undeniable that a sizable group of immigrant children and adolescents are growing up in very different circumstances than those described in the existing research literature. The context of reception provided by small towns and rural areas accustomed for decades to a uniform white/black racial makeup is very different from that encountered by migrants and their children in New York City, Los Angeles, or Miami. Theirs is truly “assimilation without a blueprint.”

Hispanic Students in New Places of Settlement

Projections from the U.S. Department of Education indicate that by 2023 Hispanic students will account for nearly 30 percent of total enrollments from early childhood through grade 12. Latino students now con-

² Field interviews conducted by the principal investigator in this town, November 2013.

stitute more than one-quarter of all K–12 enrollments. While the term “Hispanic” or “Latino” encompasses a number of nationalities, the school-level population designated by these terms is formed, overwhelmingly, by children of Mexican and Central American immigrants. Between 1995 and 2011, the Hispanic student population more than doubled in the Midwest (4.2 percent to 10.7 percent) and nearly doubled in the South (11.7 percent to 23.0 percent). Hispanic students comprise, by far, the fastest-growing segment of the school population in these regions (U.S. Department of Education, 2013).

It is not the case that the situation and educational needs of these students have been neglected either in policy or research. A wide array of educational policies and programs has been devised to address such needs, and there is a veritable mountain of reports seeking to determine which policies and school-led programs are most effective. The population of interest for such studies, as well as the relevant policies, are seldom referred to as children of immigrants or immigrant children, but rather as “ELLs” (English language learners), and the focus of attention has been on how the educational achievement of these students compare with the mainstream school population.

This focus has several notable limitations. First, it centers exclusively on *linguistic* differences to the neglect of other factors affecting children’s achievement such as family socio-economic status, family composition, parental human capital, length of U.S. residence, and neighborhood characteristics including class and ethnic composition. The sociological literature on second-generation adaptation that highlights the importance of these factors has been, by and large, ignored in policies designed to deal with ELLs. Second, all students so designated within a school or school district tend to be dealt with in a uniform manner, neglecting differences in birthplace, length of residence in the country and local community, national origin, and the family and neighborhood characteristics just mentioned. Third, who is or is not an ELL depends on scores on tests that vary significantly from district to district; even if the same test is applied, the cut-off points may be different, leading to a situation in which the same student may be placed in a remedial program in some schools, but not in others.

Policies and Practices

To address these issues, it is necessary to determine what policies and practices are being applied to immigrant students at present and what is the consensus, if any, concerning the effectiveness.

The Elementary and Secondary Education Act (1965) provided funding specifically for bilingual education. This Act was overridden by Title

III of the No Child Left Behind Act (NCLB) of 2001 that replaced the standard for bilingual education with English-only instruction and English language immersion. At the same time, the Individuals with Disabilities Education Act (IDEA) of 1975 guarantees a “free and appropriate public education for students with disabilities.” In this context, ELLs are defined as suffering from such “disabilities” and are, therefore, eligible for remedial programs. Accordingly, the ratio of Hispanic students enrolled in IDEA programs has outpaced their overall population growth. From 2000 to 2012, Hispanic students participating in IDEA programs grew by 51.8 percent, while white and black student participation declined (U.S. Department of Education, 2013). At the present, four major types of programs are designed to address the needs of ELLs. For the most part, these programs are based on some form of language immersion aimed at separating children from their linguistic and cultural background and accelerating their transition into English.

A substantial empirical literature has developed around the topic of bilingual education. For the most part, it converges on the notion that learning content and developing skills *in a child's native language* contributes to faster and better acquisition of English. According to Kenji Hakuta, arguably the most prominent exponent of this position, incorporating a student's first language into the classroom learning environment is an effective way of promoting the child's intellectual development. While some scholars have expressed concern over the reduced exposure to English in bilingual settings, a review of 15 methodologically sound studies by the National Literacy Panel found that:

Children in bilingual programs not only developed facility with English literacy to the same extent as their peers taught in English, but developed literary skills in their native language. Thus, they achieved the advantage of being bilingual and biliterate. (August et al., 2008, p. 140)

Despite this and other evidence in favor of dual language programs, the passage of NCLB led to a rapid shift in favor of structured English immersion programs and the re-definition of foreign language-speaking students as suffering from a learning “disability.” According to August and Hakuta, these transition programs are not based on research, but rather have relied on professional intuitions, political voices, and a moral conviction that “something had to be done to reverse a pattern of poor academic performance for these students” (National Research Council, 1997).

Despite some differences, the general consensus in the research literature is that none of the existing English immersion programs has succeeded in erasing the significant academic handicap of ELL students in general and Hispanic ones, in particular. According to the most recent

reports, only 22 percent of eighth-grade Hispanic students scored in the "at or above proficient" category in reading compared to 46 percent of white students. Only 26 percent of Hispanic fourth graders scored in this category for math, relative to 54 percent among their white peers (National Council of La Raza, 2015, p. 8).

These conclusions are subject to several qualifications. First, the label "Hispanic" or "Latino" is too imprecise, mixing, in unknown proportions, U.S.-born and foreign-born children and those from many different national origins. Second, the label "ELL" is also too imprecise as a designation for children of immigrants because it excludes those who, for one reason or another, are sufficiently proficient in English to escape this classification. The educational performance of such students is seldom included in existing evaluations. Nevertheless, it is clear that the massive arrival of poor and unskilled migrants in new areas of the country and the enrollment of their children in schools that, for the most part, disregard their cultural and linguistic background is leading to the rapid rise of a new caste-like population at the bottom of society. Instead of No Child Left Behind, we now have millions who are actually falling behind or dropping out of school.

What Is to Be Done?

The application of NCLB policies and their requirements have led to an increasing homogenization of educational practices and performance assessments nation-wide. The crisis being created by the application of English immersion programs is particularly acute in the states of the Midwest and South, unaccustomed to the presence of a large Hispanic population and where migrant families from that background lack the cultural support and social anchoring necessary to compensate for what is being done to their children in school. The crisis is silent because this population is still young and because the migrant communities to which they belong are still recent and unsettled. However, all signs are that we are marching headlong into an explosive situation, already adumbrated by the rapid growth of incarcerated youths of Hispanic origin in these regions.

That situation may entail the extension of the inner-city slum drama that has long been the blight of the nation to small towns and rural areas in states where new immigrants concentrate. Preventing such outcomes from becoming reality would require an end to the erroneous educational policies now in place; the implementation of true bilingual instruction programs available to both migrant and native students; and the mobilization of immigrant families and community leaders in support of keeping children in school and heightening their academic achievement and

motivation. Forcefully suppressing the culture of immigrants has seldom worked and has often backfired. What we are witnessing today is a textbook example of that outcome.

References

- August, D., Beck, I. L., Calderon, N., Francis, D. J., Lesaux, N. K., & Shanahan, T. (2008). Instruction and professional development. In D. August & T. Shanahan (Eds.), *Developing reading and writing in second language learners* (pp. 131–250). Mahwah, NJ: Lawrence Erlbaum.
- National Council of La Raza. (2015). *Latinos in new spaces: Emerging trends and implications for federal education policy*. Washington, DC: Author.
- National Research Council. (1997). *Improving schooling for language minority children: A research agenda*. D. August & K. Hakuta (Eds.). Washington, DC: National Academy Press.
- Portes, A., & Rumbaut, R. G. (2014). *Immigrant America*. (4th ed.). Berkeley, CA: University of California Press.
- U.S. Department of Education. (2013). *Annual Report*. Washington, DC: U.S. Department of Education.

Anniversary Reflections as Opportunity Both for Accruing Insights and Acknowledging Needs

Margaret Beale Spencer¹

An important observation of 50 years ago, which continues today and must be acknowledged as part of the National Academy of Education (NAEd) 50th Anniversary Celebration, is the continuing salience of race in American education. The overlapping themes having to do with the unique and interacting influences of *identifiability, race, and group membership continue as part of the fabric of education, research, policy, and practice in the 21st century*. The continuing American dilemma for the nation and professional organizations suggests an inability to acknowledge and cope with human differences (i.e., both informally as everyday practices and formally as social science research traditions). Both situations matter equally, deeply and profoundly. What is critical about the NAEd today is that there are individual and organizational efforts to negate the noted untoward traditions and *to press against the uncomfortable acknowledgment that race and other indicators of group membership matter as continuing 21st-century America challenges*. Tate (1997) described the dilemma quite cogently when reporting on the controversy surrounding then University of Pennsylvania law professor Lani Guinier's research when nominated for the position of assistant attorney general in charge of the Civil Rights Division. He noted:

¹ Margaret Beale Spencer is the Marshall Field IV Professor of Urban Education and Professor of Human Development at the University of Chicago. She was elected to the National Academy of Education in 2009.

Guinier's research, which examined voting systems, asked the following question: Are there factors that guarantee winners and losers? She contended that such factors do exist and that race is too often an important factor in the construction of voting districts, the outcome of elections, and ultimately political influence, including the control of educational systems. (Tate, 1997, p. 195)

He continued to underscore the everyday salience of race by noting: "Ladson-Billings and Tate (1995) asserted that, despite the salience of race in U.S. Society, it remains untheorized as a topic of scholarly inquiry in education" (cited from Tate, 1997, p. 196). *The point stressed here is that the fact of socially constructed parameters of human difference and its particular manifestation as race continues to matter profoundly in the social sciences and appears most deleterious in regards education as opportunity.*

As indicated by education legislation passed beginning in the mid-1950s, too frequently the school as the context for learning and development—given its educational mission—is a pivotal social system. It represents the site of traditions, which obstruct the actual potential of education (i.e., determining the fact of life course "winners and losers" as suggested by Guinier and noted by Tate, 1997); in other words, the content, quality, and character of educational contexts serve as the vehicle of life course opportunity (i.e., manifested as "winners and losers"). The race-associated, relationally based subtleties coped with by youth serve to inhibit (or, on the other hand, to privilege) opportunity for life course learning and development and, without question, continue uninhibited today as gap findings. The strengths implicit in interdisciplinary-framed education research efforts should aid in moving education practices and policies from a focus on the "what" as gap outcomes (i.e., as assisted by innovative statistical insights) to an appreciation of the "how" and "why," which afford the needed specificity for change (i.e., as training, practice, and policy) as well as improved use of government funding.

Fifty years ago and following *Brown v. Board of Education* (1954, 1955) education legislation, unfortunately, the proclivity to ignore the impact of race continues as the NAEd prepares to celebrate its 50th anniversary, and the nation remains reluctant to acknowledge the significance and need for the recently formed "Black Lives Matter" movement. That is, the continuing salience of race in all American contexts "post-passage" of desegregation education legislation (i.e., as *Brown v. Board of Education* [1954, 1955])—and embarrassingly so—endures as an uncomfortable social reality. Importantly, the fact is not lost on the unavoidable cognitive constructions of social life inferred (and acted upon) by American youth as normative privileges for some and everyday ignored challenges for others.

The major point for my particular reflections is that as we celebrate the anniversary of the organization, it is critical to acknowledge that—as associated with race, ethnicity, gender, and the intersectionality of socio-economic status with the previous—*there continues to be life-changing work to accomplish for all Americans*. That fact is the reality irrespective of efforts proffered to date. Deficit-based underlying assumptions about group differences continue to hold sway. More specifically and although more sophisticated in character, conceptually biased questions, “context free approaches,” and methodologically limited research designs are reported as frequently today as during the five intermittent decades that heralded the founding of the organization.

From my perspective, what is hopeful today over the situation 50 years ago is that the NAEd—particularly given its links with relatively diverse organizations such as the American Educational Research Association—represents opportunity. I infer a conviction of some members to acknowledge privilege and the scholarly representation of myriad socially constructed “human differences.” My sense is that there is an interest from some members to have the organization and social sciences, more generally, to be perceived in the future on the more “positive side of history.” I make the statement as a potential, which is not necessarily a given and anticipated reality unless conceptualizations about the meaning of “social differences” shift; also needed will be proactively evolved beliefs concerning equity vis-à-vis inclusion and under-acknowledged conditions of privilege (i.e., relevant to both education and justice systems and which have stress-linked health implications).

The latter themes are critical. When considering the conceptual contributions of privilege and critical race perspectives to the social sciences, too *infrequently acknowledged* is that each uniquely provides substantive understandings about learning and human development processes across the life course. Along with normal human development theorizing, the noted frameworks contribute insights about the variable levels of vulnerability of humans given socially constructed differences in access to opportunity.

The NAEd is particularly well suited to tackle the conceptual challenges given the interdisciplinary character of its membership. The noted conceptual frames afford insights about perceived coping needs and adaptive processes required of diverse citizens for maximizing learning, development, and sought after contributions to the broader society as providers to the national commonwealth; this is different from narrow or limited views held about particular groups given assumptions of a limited social status as “recipients.” Additionally, as suggested previously, each perspective contributes intuitions about patterned life course successes as well as social and contextual conditions that precipitate persistent chal-

lenges, thus informing sources of between- and within-group variations as well as incongruent outcomes. Redundant media-disseminated education gap findings—which too frequently contribute to stereotyping—frame life course group statuses (i.e., as “cradle to coffin” media-reported differences). Stigmatized and dissimilar relational and contextual processes and outcomes contribute to under-acknowledged individual-context experiences. Unique NAEd workgroups, which represent across-university collaborations, have great potential for shifting the level of discussion and altering the character of science and constructs included in the conduct of science. *Thus—as more enlightened views concerning group membership candidly discussed, potentially through the NAEd collaborations—privilege and critical race perspectives when linked with non-pathologizing developmental science perspectives afford new conceptual possibilities. Their exploration become more probable scholarly frames independent of the prevailing Zeitgeist at one’s home institution.* As noted, the recent emergence, visibility, and cross-race instigation of the “Black Lives Matter” movement has drawn attention to the social dilemma of race-linked inequality; the view is as important for social life as for scholarly efforts. Thus, our consideration of its impact should be nuanced efforts, which also afford and focus attention on myriad settings where learning and development occur. My point is that diverse citizens—children and adults—navigate environments and attempt access to socially constructed supports, which are constitutionally promised but, in fact, are blocked by inequitable access. Accordingly, the situation requires theoretically driven, conceptually nuanced, and methodologically enlightened strategies, which may require scholarly collaborations that go beyond one’s immediate institutional options and proximal supports. This celebratory occasion—in moving forward—can acknowledge the organization’s potential for affording the type of needed vehicle noted.

To summarize, the NAEd is celebrating its 50th anniversary; however, race has continued to be treated in the social sciences and education literatures in rather short-sighted ways for a period exceeding 75 years. However, I am hopeful because the NAEd’s multi-disciplinary member constituency provides opportunities for unique collaborations for impacting how “difference” is conceptualized and operationalized through sensitive and reality-based cooperative education research, policy, and practice informing efforts.

References

- McGee, E. O., & Spencer, M. B. (2013). The development of coping skills for science, technology, engineering, and mathematics students: Transitioning from minority to majority environments. In C. C. Yeakey, V. S. Thompson, & A. Wells (Eds.), *Urban ills: Post recession complexities of urban living in the twenty first century* (pp. 351–378). Lanham, MD: Lexington Books.
- Tate, W. F. (1997). Critical race theory and education: History, theory, and implications. *Review of Research in Education*, 22, 195–247.

Assessment and Measurement in Education

Assessment: Friend or Foe of Pedagogy and Learning

*Paul Black*¹

The long-established theories of pedagogy pay only marginal attention to assessment (Bruner, 1999): it is merely an activity at the end to evaluate outcomes. For many teachers, assessment is the unpleasant part of their work—a view strengthened by the pressures of high-stakes external testing. Yet the growth of formative assessment has altered this perspective for many, for whom the tension is now between formative assessment Good and summative assessment Bad. To resolve this tension, we need a new model of pedagogy in which assessment is seen to have a central role. I will sketch out such a model and show how it incorporates a unified view of assessment, one that reconciles the formative and the summative while justifying the alternative name—Assessment for Learning (AfL).

Assessment and Pedagogy

Any model of pedagogy should have the development of students' learning as its central aim. I argue that this implies that assessments, ranging from the formative to the summative, must be integrated within the pedagogical structure. To support this assertion, I will draw upon studies in which our group at King's College explored how results from research on formative assessment could be used to improve classroom teaching.

¹ Paul Black is Emeritus Professor of Science Education at King's College, London. He was elected to the National Academy of Education in 2010.

These studies started from the following passage in Black and Wiliam's 1998 review:

What does emerge is a set of guiding principles, with the general caveat that the changes in classroom practice that are needed are central rather than marginal, and have to be incorporated by each teacher into his or her practice in his or her own way. (p. 62)

If you judge by the number of citations, the review by Black and Wiliam (1998a), together with the brief summary of the main conclusions that they also published (1998b, 1998c), had a big impact. This article assembled a wide range of evidence that supported the belief that formative feedback improves learning. However, this wide range may have led to a diversity of implications for practice, so its impact has been problematic (Bennett, 2011).

To explore this caveat, my colleagues and I set up a 2-year project with 40 teachers from six schools (Black, Harrison, Lee, Marshall, & Wiliam, 2003). While explaining to the teachers the various findings in the review, we emphasized that they would have to transform the methods explored in the research into methods that would work in their classrooms. For the first 9 months, the teachers tried different methods from the research findings and met regularly to share their experiences. In a second phase, each teacher chose the methods that seemed most useful for him or her and applied the methods consistently within one "target" class. One outcome was that the target classes produced better test scores than did comparable classes (Wiliam, Lee, Harrison, & Black, 2004). Another was that the teachers used widely varying sets of methods. The teachers' written reflections were recorded in a book about the project (Black, Harrison, Lee, Marshall, & Wiliam, 2003); I use extracts from the book in this article.

In reflecting on this project, my colleagues and I realized the need to broaden our perspective, which was confirmed in a 1998 review of our work by Perrenoud, who said that our focus on feedback was too narrow:

I would like to suggest several ways forward, based on distinguishing two levels of the management of situations which favour the interactive regulation of learning processes:

- the first relates to the setting up of such situations through much larger mechanisms and classroom management
- the second relates to interactive regulation which takes place through didactic situations. (Perrenoud, 1998, p. 92)

The Five Stages of AfL

In presenting the following argument about assessment in pedagogy, I take Perrenoud's comment seriously by proposing the following model, which incorporates AfL in each of its five stages:

- A. Decide learning aims
- B. Select and plan activities to achieve those aims
- C. Implement in the classroom
- D. Review: informal summative assessment
- E. Formal summative assessment

In stage A, aims are often expressed in such vague terms that only in high-stakes formal assessment are their meanings made clear. Ideally, there should be synergy between stage A and stage E. The assessments used in stage E should be designed alongside the formulation of aims in stage A, so that they are not after-thoughts but are borne in mind at all stages and thereby serve to reinforce the aims (Klenowski & Wyatt-Smith, 2014).

Perrenoud's first and second levels correspond to stages B and C. For stage B, the task is to plan activities in ways that both engage students' participation and promote their learning. To develop this point, I provide a quote from a teacher who prepared for a first lesson on photosynthesis by keeping two pots of the same plant in different locations: in a dark corner and by a window. After a few weeks, he presented the plants to his students and asked them to discuss with one another why the plants looked so different. He then invited the students to exchange their ideas, which started as follows:

T: Monica—your group? Pair?

Monica: That one's grown bigger because it was on the window.

T: On the window? Mmm. What do you think Jamie?

Jamie: We thought that. . .

T: You thought. . .?

Jamie: That the big'un had eaten up more light.

T: I think I know what Monica and Jamie are getting at, but can anyone put the ideas together? Window - Light - Plants?

Richard: Err.... We thought, me and Dean, that it had grown bigger because it was getting more food. (Black et al., 2003, p. 38)

This example illustrates two key features. First, the teacher used formative assessment to generate information to be used as feedback. From the students' responses, the teacher learned about their prior understandings of the topic, so that he could guide them to develop their ideas from

these starting points. At the same time, the teacher encouraged the students to engage in a learning dialogue, strengthened by feedback among them. Alexander (2006) emphasized the importance of learning dialogue:

Children, we now know, need to talk, and to experience a rich diet of spoken language, in order to think and to learn. Reading, writing and number may be acknowledged curriculum “basics,” but talk is arguably the true foundation of learning. (p. 9)

Similarly from Wood (1998):

Such encounters are the source of experiences which eventually create the ‘inner dialogues’ that form the process of mental self-regulation. Viewed in this way, learning is taking place on at least two levels: the child is learning about the task, developing “local expertise”; and he is also learning how to structure his own learning and reasoning. (p. 98)

Stages B and C interact: a task designed in stage B may only work if the teacher can achieve a balance between encouraging a range of students’ inputs and steering a course toward the learning aim. This balance was described by Black and Wiliam (2009), who identified two key elements of formative assessment as “engineering effective classroom discussions” and “providing feedback that moves all learners forward.” (p. 8)

Unfortunately, evidence from surveys of classrooms in the United Kingdom and the United States shows that the quality of dialogue is poor, characterized, in Alexander’s terms, as involving far more of the rote-learning and instruction/exposition styles than of the discussion and dialogue styles (Alexander, 2006; see also Applebee, Langer, Nystrand, & Gamoran, 2003).

In addition to verbal dialogue, there is also dialogue as students write homework or seat-work or answer questions in a short test. This occurs during stage D, when students review and express what they learned and thereby enhance their learning. The teacher may then give individual feedback. However, the way in which feedback is given is crucial. As Dweck’s research has shown, students provided feedback in the form of marks are likely to view the feedback as a way to compare themselves with others and students provided feedback in the form of comments as a way to help them to improve: the latter group outperforms the former (Dweck, 2000). Many teachers in our project stopped grading students’ work, focusing instead on using comments to advise each student on how to improve his or her work. One teacher reported her experience as follows:

(The researcher) observed on several occasions how little time students spend reading my comments if there were grades given as well. My routine is now, in my target class, to: (i) not give grades only comments;

(ii) comments highlight what has been done well and what needs further work; (iii) the minimum follow-up work expected to be completed next time I read the books. (Black et al., 2003, pp. 43–44)

Within the same project, students marked one another's work. One teacher described this experience as follows:

We regularly do peer marking—I find this very helpful indeed. A lot of misconceptions come to the fore and we then discuss these as we are going over the homework. I then go over the peer marking and talk to pupils individually as I go round the room. (Black et al., 2003, pp. 50–51)

Some teachers took peer assessment even further by using it in marking informal, end-of-topic tests. One teacher described her students' experience of this approach as follows:

[T]hey are very positive about the effects. Some of their comments show that they are starting to value the learning process more highly and they appreciate the fact that misunderstandings are given time to be resolved, either in groups or by me. They feel that the pressure to succeed in tests is being replaced by the need to understand the work that has been covered and the test is just an assessment along the way of what needs more work and what seems to be fine. (Black et al., 2003, pp. 56–57)

For such discussion to work, students must be able to collaborate effectively in group discussion. The review of Johnson, Johnson, & Stanne (2000) showed that this ability cannot be taken for granted. Mercer, Dawes, Wegerif, & Sams (2004) developed and tested a program in which students were trained in group behavior. This training improved the students' subsequent test scores, but it also revealed that the students used such terms as "because," "think," "should," "would," three times more often after the training.

In summary the work in stage D, through its engagement of students in peer assessments, all in the light of the overall aims, helps students develop an overview of their learning. Such work implements two more learning strategies: activating students as instructional resources for one another, and activating students as the owners of their own learning (William & Thompson, 2007).

Overall, the learning habits that students may develop through the approaches described above may be summarized as follows:

1. engage in and learn from interactive dialogue: stage C
2. reflect on critical scrutiny of their work, and use this to guide improvement: stage D
3. collaborate in group-work: stages C and D

4. achieve a broad overview of their progress and guide their development in the light of the aims of the learning: stage D

This leaves Stage E. The priority here should be to guide decisions about each student's future. Where external high-stakes tests are involved, teachers may be victims of systems with narrow aims. However, many tests within schools do not fall into this category, for example, those at the end of a semester that guide decisions about students' future study. A small-scale study in the United Kingdom of teachers' summative assessments showed that they did not have clear policies for their own summative tests (Black, Harrison, Hodgen, Marshall, & Serret, 2011). They never studied the concepts as validity and reliability, and, lacking confidence in their ability to set formal tests, they copied items from the Internet that did not reflect their learning aims. Two years of work with 30 teachers achieved significant improvements, both in their practices and their confidence. The work included both the broadening of assessment methods beyond the constraints of formal testing and the use of inter-school and intra-school "moderation" meetings in which teachers compared samples of their students' work, with their grading judgments, to ensure alignment of standards. Those meetings achieved far more than simply checking grades. As one teacher of English described,

I think it's quite a healthy thing for a department to be doing because I think it will encourage people to have conversations and it's about teaching and learning.... It really provides a discussion hopefully as well to talk about quality and you know what you think of was a success in English. Still really fundamental conversations. (Black, Harrison, Hodgen, Marshall, & Serret, 2011, p. 461)

The significant point here is that when teachers build confidence and trust, the work of stage E can help them to achieve their overall learning aims. The pressures of high-stakes accountability testing may not go away, but they may directly affect work only in those testing years—schools could ignore the pressures in other school years.

The Impact of AfL

Given the claims made above, it is natural to ask why AfL's impact has been uneven within the past 15 years. The evidence that it can lead to improved test scores is, at best, patchy (Kingston & Nash, 2011), and there is evidence of wide variations in the practices that claim to implement the AfL principles. These problems were anticipated in the following caveats by Black and Wiliam (1998b, 1998c):

Thus the improvement of formative assessment cannot be a simple matter. There is no “quick fix” that can be added to existing practice with promise of rapid reward. On the contrary, if the substantial rewards of which the evidence holds out promise are to be secured, this will only come about if each teacher finds his or her own ways of incorporating the lessons and ideas that are set out above into his or her own patterns of classroom work. This can only happen relatively slowly, and through sustained programmes of professional development and support. (1998b, p. 15)

Given these reservations, and the lack of “sustained programmes of professional development,” the dearth of evidence about implementation of AfL may have been expected.

On a more positive note, the following reflection by a deputy principal from a school involved in our AfL project summarizes my main message:

[It's essential] that we have a greater emphasis on children's learning, that we are supporting learning far more than we are doing at the moment. I don't think that we do it particularly well. Individuals do but I don't think that we are using our assessment to progress learning. It doesn't happen overnight. So if you are saying “what do I want in five years time?”—ideally it's that all staff are using assessment as a tool to develop children's learning. (Black et al., 2003, p. 103)

References

- Alexander, R. (2006). *Towards dialogic thinking: Rethinking classroom talk*. York, UK: Dialogos.
- Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal*, 40(3), 685–730.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education*, 18(1), 5–25.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1).
- Black, P., & Wiliam, D. (1998b). *Inside the black box: Raising standards through classroom assessment*. London, UK: GL Assessment.
- Black, P., & Wiliam, D. (1998c). Inside the black box: Raising standards through classroom assessment. *PhiDeltaKappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning—putting it into practice*. Buckingham, UK: Open University Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, M., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education*, 18(4), 451–469.
- Bruner, J. S. (1999). Folk pedagogies. In J. Leach and B. Moon (Eds.). *Learners and pedagogy*. London, UK: Paul Chapman.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Psychology Press.

- Johnson, D. W., Johnson, R. T., & Stanne, M. B. (2000). *Cooperative learning methods: A meta-analysis*. Minneapolis, MN: University of Minnesota.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
- Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: A guide for students, teachers and researchers*. London, UK: Sage, p. 105.
- Mercer, N., Dawes, L., Wegerif, R., & Sams, C. (2004). Reasoning as a scientist: ways of helping children to use language to learn science. *British Educational Research Journal*, 30(3), 359–377.
- Perrenoud, P. (1998). From formative evaluation to a controlled regulation of learning processes. Towards a wider conceptual field. *Assessment in Education: Principles, Policy and Practice*, 5(1), 85–102.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.
- Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education*, 11(1), 49–65.
- Wood, D. (1998). *How children think and learn* (2nd ed.). Oxford, UK: Blackwells.

Conservation and Exploration: What Is, What If, Why Not?

*Edmund W. Gordon*¹

I enjoy welcoming newly elected National Academy of Education (NAEd) members with a perspective on what I think it means to be a member of the National Academy of Education, where I have enjoyed the privilege of membership for nearly half a century. When I was invited into membership, my first response was a feeling of appreciation for having been recognized and for being honored by the invitation. That appreciation was soon coupled with a sense of responsibility for the conservation and advancement of the conceptual richness for which the work of the NAEd members is known. Association with my colleagues in the Academy constantly reminded me of the expectation and even obligation to maintain that high standard in my own work and to preserve and protect not only the richness, but also the authority that flows from that tradition. Less ubiquitous, but equally compelling, was evidence of a second value prevalent among some of my colleagues: a sense of responsibility for intellectual leadership to parallel the responsibility for conservation. Along with the protection of the best that is, they assumed the responsibility to maintain room, even make space, for that which is new. That is, my esteemed colleagues at the Academy refuse to be so constrained by what we know as to render us unreceptive to that which may be possible. Like

¹ Edmund W. Gordon is the Richard March Hoe Professor Emeritus of Psychology and Education at Yale University and Teachers College, Columbia University. He was elected to the National Academy of Education in 1979.

C. Wright Mills, we ask “what if” questions, and if there is reason to do so, like Robert Kennedy, we ask “Why not?”

At the beginning of my tenth decade of life, I found myself confronted with a unique challenge to honor and hold in balance these two values—conservation and informed exploration. I was invited to serve as chairperson of the Gordon Commission on the Future of Assessment in Education (Gordon Commission). In that role, I was called upon to respect and protect a well-developed science and technology of measurement, even while criticizing it for its failure to keep pace with the demands of a changing society. At the same time, I was becoming increasingly aware that the measurement enterprise has the capability to move beyond its excellence in the documenting of status toward enabling us to understand the processes by which status is achieved, and even to contribute to its achievement. My colleagues on the Commission led me to believe that the conjoining of assessment and education can certainly contribute to accountability, but it can also contribute to our understanding, informing, and improving of teaching and learning processes and their outcomes.

In education, we have long struggled to rise to the challenge of enabling the full development of all persons in our society, given that there is such great variation in the manifestations of traits and conditions in the human population. We now face a second great challenge, which is that of responding to the global challenge of demands and opportunities presented by the digital revolution in the production of commodities, knowledge, relationships, and even intellect itself. In navigating this second challenge, education may be not only the quintessentially human resource, but also the essential as well. I use the term “education” here in the sense that Martinez referred to the process as “the cultivation of intelligence.” In another context I have referred to education as a dialectical pedagogical troika, in which assessment, teaching (inclusive of curriculum), and learning are integrated in the service of human development.

It is within this conceptual context that I turn to the problems of how, given the diversity in human conditions and characteristics, we may achieve equity (as opposed to equality) in the opportunity to develop human intellectual capacities, and excellence in the quality with which intellectual competence is achieved.

Building on the work of the Gordon Commission, my colleagues and I have undertaken a conceptual inquiry into whether measurement science can in fact improve the processes and outcomes of teaching and learning in the science, technology, engineering, and mathematics (STEM) disciplines. These inquiries have stimulated me to go further and to ask the question of whether assessment can be educative and actually help to cultivate intellect in addition to its accomplishments in measuring developed abilities. It is my sense that the excellence that has been associ-

ated with the scholarship in this domain suggests that this science can be effectively applied to the challenge of capacity building by:

1. Analyzing the processes by which intellectual competence is developed through learning and teaching rather than the more traditional emphasis on the measurement of the status of developed abilities that have or have not resulted from such processes. Assessment can be focused on developmental process in addition to focusing on developmental product.
2. Appraisal of learners' access to and utilization of resources as a diagnostic tool to inform pedagogical intervention.
3. Using assessment probes and learners' responses as didactic instruments to instruct as well as to inquire.

In this work, I advance a vision for unlocking the potential of educational assessment as an integral component of pedagogy, which can be applied toward enabling all students—regardless of race, class, gender, linguistic or family background, or place of residence—to become intellectually competent and meet the challenges of the 21st century. I call this a vision of assessment FOR rather than OF education.

Many thinkers in various fields have warned us that this next century will bring unprecedented transformations to our society. We are already feeling the effects of what the marketplace expects educated citizens to be able to do, and we are beginning to see how technology can enhance or modify human abilities. Despite a steady program of educational reform, and despite concerted effort by passionate and committed educators to understand and improve pedagogy throughout the 20th century, we have failed to educate all learners in our country. The growing and ever more costly assessment regime we have designed has not only failed to tell us why we have failed in our strenuous efforts but also does not give us the information and tools we need to improve education for all learners. Assessment FOR rather than OF education may allow us to right our course as we move forward into an uncertain future. Obviously, the field of measurement science is capable of doing far more than is reflected in current policy and practice. Our task is to preserve the best of what we know how to do in measurement, even while we actively pursue new challenges and opportunities.

Refining Assessment to Strengthen Online Science Learning

Marcia C. Linn^{1,2}

The National Academy of Education recommended use of technology for science instruction, informed by research over the past 50 years (National Academy of Education, 2009). Recent reviews of research on scientific visualizations (McElhaney, Chang, Chiu, & Linn, 2015), guidance based on natural language processing (Gerard, Matuk, McElhaney, & Linn, 2015; VanLehn, 2011), and science learning environments (Donnelly, Linn, & Ludvigsen, 2014) reinforce this recommendation. Furthermore these technologies offer excellent opportunities to make assessments learning opportunities by logging student and teacher actions, making use of innovative response formats including essays and graph construction activities (e.g., Vitale, Lai, & Linn, 2015), and automatically scoring student work.

Yet, schools are allocating their scarce computer resources to high-stakes testing that is disconnected from learning. Some schools reassign computers purchased for instruction for practice tests and administration of the assessments for up to 60 days of the school year (Chingos, 2012; Nelson, 2013).

¹ Marcia C. Linn is Professor of Cognition and Development at the Graduate School of Education, University of California, Berkeley. She was elected to the National Academy of Education in 2007.

² This material is based on work supported by the National Science Foundation under NSF Projects 1119670, 0822388, 0918743. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

These assessments rarely meet teacher or student needs. As one teacher, calling for a moratorium on testing, remarked: “Students come to us with an enormous range of abilities.... It is a contradiction to ask us to individualize instruction and then administer a single standardized test that will be used to evaluate and rate entire school systems, as well as individual teachers and students” (Garro, 2015).

Instead, technology can increase student success in science by connecting instruction and assessment. The NAEd White Paper on Science and Mathematics Education called for “comprehensive technology-based instructional and assessment resources ... [and an] iterative cycle of research and development” (NAEd, 2009, pp. 7, 9).

Technological advances make this agenda feasible. First, delivering instruction using an online learning environment can enable randomized assignment to instructional conditions even within the same classroom to support iterative cycles of research and development. Second, personalized instruction that takes advantage of student prior knowledge can increase student agency, identity with science, and coherent understanding of complex topics (e.g., Linn & Eylon, 2011). Third, logging of student interactions with scientific models and visualizations and guiding students based on their trajectories can capture rich evidence of student learning. Fourth, real-time analysis of performance on assessments embedded in instructional materials can provide timely information for teachers, who often get standardized test results months or even a full year after the tests are administered. Teachers can use timely information to customize their instruction to student needs, identify students who are struggling, and allocate their time efficiently.

Rather than devoting up to 60 days to high-stakes assessments, science teachers could use those days, along with the computers, for instruction that takes advantage of embedded assessments to monitor progress. By building on what we have learned about technology-enhanced science instruction over the past 50 years we can connect instruction and assessment and improve outcomes. Several examples from my own experience illustrate how the iterative cycle of research and development can strengthen science teaching and learning.

Personalizing Instruction

Although we have learned a great deal about how to customize instruction for students with varied prior knowledge, cultural expectations, language experiences, and learning strategies, our assessment and instructional systems can do more to incorporate these findings. Work on aptitude-treatment-interactions initiated by Lee Cronbach and Richard Snow at Stanford (while I was a student there) illustrate how technologi-

cal environments might capitalize on the aptitudes of individual learners (Cronbach & Snow, 1977). They found, for example, that task structure interacted with student anxiety: students with high anxiety tended to succeed in structured instructional environments while low anxiety students succeeded with low structure, a variable that technology-enhanced instruction could vary.

One aptitude-treatment-interaction study had valuable implications for science instruction. Researchers investigated the advantage of matching students' spatial and verbal aptitude to two forms of economics instruction. They found that building on strengths might be less effective than remedying weaknesses. Assigning those low on spatial reasoning to struggle to learn with complex diagrams and by challenging those high on spatial reasoning to make sense of verbal descriptions of the material was more successful than matching learners and aptitudes. Effective instruction essentially added difficulties by pushing students to develop weak abilities rather than building on their cognitive strengths (Cronbach & Snow, 1977). These results suggest research questions for investigations of spatial visualizations in science courses.

Working with Piaget in Geneva and at the Lawrence Hall of Science in Berkeley introduced me to the fascinating and often contradictory ideas that students bring to science activities. These ideas (often called misconceptions) have been studied for a broad range of science topics (e.g., Driver, Guesne, & Tiberghien, 1985; Inhelder & Piaget, 1969). Research also shows that typical courses involving lectures, demonstrations, and experiments rarely modify these ideas (Halloun & Hestenes, 1985). Studies of the origin of these ideas revealed that students often base their thinking on accurate observations and promising forms of scientific reasoning (e.g., diSessa, 1988). Students develop a repertoire of fragmented and contextualized ideas such as the notion that heating is a distinct process from cooling or the idea that objects in motion remain in motion in science class, but not on the soccer field (Linn & Eylon, 2011). Encouraging students to use evidence to sort out their own ideas as they encounter new ideas can affirm the value of struggling to understand complex phenomena. Such struggles can develop a feeling of agency and identity with science, an especially crucial aspect of instruction for science students from nondominant cultures (Brown, 2004).

Specifically, the ideas students develop while exploring the world offer useful building blocks for the formation of more sophisticated views. For example, Hammer (1996) guided students to test their inaccurate ideas and found this led them to gain insight into normative ideas. He found that when students tested their predictions about why a ball rolling down a track picks up speed they struggled to explain why their predic-

tions were wrong. They gained insight into force and momentum and were then able to use these normative ideas to interpret new challenges.

Documenting how students sort out their prior knowledge and new information using a computer learning environment can inform research and development cycles. This documentation can provide a trajectory of student progress that allows teachers to assess both the final understanding of the student and the strategies the student uses to make sense of alternative views of scientific phenomena.

Taking Advantage of Scientific Visualizations

Scientific visualizations are widely used in research laboratories and offer great promise for strengthening science instruction. Online systems can enable students to interact with phenomena that cannot be observed directly because they are too small (atoms and molecules), fast (collisions), vast (solar system), or dangerous (explosions). Studies show the advantages of visualizations, the potential of dynamic depictions (Ryoo & Linn, 2012), and the challenges of making visualizations comprehensible (McElhaney, Chang, Chiu, & Linn, 2015).

Numerous studies report on how researchers test and refine visualizations embedded in inquiry instruction to ensure that they help students learn. For example, Helen Zhang (Zhang & Linn, 2011) conducted a series of iterative refinements of a Web-based Inquiry Science Environment (WISE) unit featuring a Molecular Workbench (Xie & Tinker, 2006) visualization. In the unit, students learn about chemical reactions by exploring how the combustion of hydrogen and oxygen provide energy for a hydrogen fuel cell car. Zhang led a partnership of teachers, technologists, chemists, and researchers to design the unit.

They documented that students often gain superficial understanding from visualizations of chemical reactions by asking students to use a drawing tool to depict the reaction before it started, when it started, at the midpoint, and when it completed. Student drawings showed that many assumed that the chemical equation told the whole story. Students often drew the molecules on the two sides of the equation as the start and completion of the reaction and left the two middle spaces blank. Students neglected the process of bond breaking and bond formation, failed to conserve matter, added conditions that did not occur such as breaking all the molecules into atoms, and created impossible configurations of atoms.

Based on the assessments, Zhang integrated the drawing tool into the instruction. As they explored the visualization, students recorded their observations in drawings. She found that in the drawing condition students struggled to interpret the visualization, often running it again and again to record bond breaking and bond formation. Drawing was more

effective than an experimentation condition where students also ran the visualization multiple times (Zhang & Linn, 2011).

The embedded assessments involving drawings and explanations helped teachers analyze their lessons and grade their students. The teachers used the insights they gained from reviewing the drawings to refine their classroom instruction (Zertuche, Gerard, & Linn, 2012).

In summary, researchers are finding ways to engage students in learning from scientific visualizations. Sustained exploration of complex visualizations can improve understanding of complex, often invisible processes and as a result, strengthen students' identity as a science learner.

Designing Adaptive Guidance

Computers can guide science students rather than primarily correcting their work by analyzing logs of activities; diagnosing flaws in drawings, graphs, geometry proofs, or concept maps; and analyzing short essays using natural language processing software (Linn et al., 2014). For example, studies have focused on automating knowledge integration guidance that has proved helpful in prior research (Linn & Eylon, 2011). Knowledge integration guidance involves acknowledging student progress, identifying an area for improvement, suggesting an action to get additional evidence (such as revisiting a visualization), supporting the student to sort out the differences between the evidence and their response (such as in the draw condition for chemical reactions), and asking the student to revise their response.

To create automated knowledge integration guidance, researchers score more than 1,000 essays written by diverse students. It would take 5 or more years for a teacher to encounter this many responses, giving the guidance program an advantage. These scores are used to train the natural language processing software developed at the Educational Testing Service (Liu et al., 2014). The WISE unit delivers guidance based on the scores awarded by the software.

A series of studies explored automated knowledge integration guidance for short essays and drawings about complex science topics including photosynthesis, mitosis, chemical reactions, and cellular respiration (Gerard, Ryoo, McElhaney, Liu, Rafferty, & Linn, 2015). The studies compared automated knowledge integration guidance to teacher-selected knowledge integration guidance; individualized guidance designed by teachers who participated in professional development, inspired successful studies such as Herrenkohl, Tasker, & White, (2011); guidance pointing out specific errors often used by automated tutors (VanLehn, 2011); and generic guidance often used when teachers lack time to construct individualized guidance (Ruiz-Primo & Furtak, 2007). Across all

experiments, automated knowledge integration guidance was as effective as individualized teacher guidance and more effective than generic or specific guidance. Furthermore, students who revisited the visualization as recommended by the guidance were more successful than those who did not revisit.

Automated guidance frees science teachers who might otherwise devote two or more hours to writing comments on the essays of their 150 students. Teachers can instead assist struggling students and use summaries of the scores to pace instruction to the needs of their students. Teachers appreciated the automated guidance. They noted that the computer guided all the students who were floundering, ensuring that even students who were reluctant to ask for help often made progress.

Because knowledge integration guidance does not provide the right answer, it challenges students to make sense of evidence that may seem contradictory or irrelevant at first. Successful struggles can prepare students to persist when their ideas seem superficial or inaccurate and promote agency. The value of these struggles is supported by research, primarily on memory tasks, demonstrating that making certain types of mistakes may increase failure during learning yet lead to better long-term retention (e.g., Richland, Linn, & Bjork, 2007; Soderstrom & Bjork, 2015). Bjork labeled these experiences as desirable difficulties. One desirable difficulty he identified is spacing rather than massing study opportunities. This is consistent with the success of revisiting evidence following knowledge integration guidance. More research is needed to understand how best to support students as they encounter challenging science topics.

Conclusions

Technological environments such as WISE can improve science curricula by incorporating visualizations, combining them with adaptive guidance, and using them to illustrate contemporary scientific issues such as global climate change and energy efficiency. These materials can promote agency and identify with science. They can prepare citizens to continuously strengthen their scientific understanding. By focusing on ongoing dilemmas science instruction can increase the likelihood that students will revisit the topics they studied in school as they reappear throughout their lives.

To be sure, improving science instruction by incorporating technology involves overcoming many obstacles. A crucial obstacle is state and national standards that mandate covering far too many topics each year. Taking advantage of cycles of research and development could help set realistic goals.

Another obstacle concerns high-stakes tests that measure recall of

details rather than argument construction and knowledge integration. These assessments not only frustrate teachers as noted in the quote from the teacher in Massachusetts but also illustrate the futility of efforts to promote knowledge integration. Science tests often have fewer items than there are topic standards for the current year!

A final obstacle is funding for technology. Schools need budgets that provide not only computers but also the technical support essential to making the computers useful. Declining costs for personal tablets and laptops is making technology more available and reducing the disparity in access. More must be done to ensure that all students, and especially those who lack access at home, have full access to the technologies that enable them to succeed.

Technology has the potential to implement effective continuous assessment and to achieve the goals outlined in the NAEd (2009) white paper. Research over the past 50 years offers excellent evidence for strengthening technologies for science learning. Rather than interrupting science instruction to allocate computers to testing, teachers can capitalize on computers to coherently improve science teaching, assessment, and learning. Achieving this vision requires concerted research that addresses the links between curriculum, assessment, student characteristics, teaching, teacher preparation, and policy. It requires overcoming many obstacles. It especially requires policy leadership that is the hallmark of the NAEd.

References

- Brown, B. A. (2004). Discursive identity: Assimilation into the culture of science and its implications for minority students. *Journal of Research in Science Teaching*, 41(8), 810–834.
- Chingos, M. M. (2012). *Strength in numbers: State spending on K–12 assessment systems*. Washington, DC: Brown Center on Education Policy, Brookings Institution. Retrieved from <http://tinyurl.com/br63uhv>.
- Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. Pufall (Eds.), *Constructivism in the computer age* (pp. 49–70). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Donnelly, D. F., Linn, M. C., & Ludvigsen, S. (2014). Impacts and characteristics of computer-based science inquiry learning environments for precollege students. *Review of Educational Research*, 20(10), 1–37. doi: 10.3102/0034654314546954.
- Driver, R., Guesne, E., & Tiberghien, A. (Eds.). (1985). *Children's ideas in science*. Philadelphia, PA: Open University Press.
- Garro, G. (2015, July 23). The argument: Should the state put a three-year moratorium on high stakes testing? *Boston Globe*. Retrieved from <https://www.bostonglobe.com/metro/regionals/north/2015/07/23/the-argument-should-state-put-three-year-moratorium-high-stakes-testing/RnJxWkdm9aSAKKHcrX3ZtK/story.html>.
- Gerard, L. F., Matuk, C. F., McElhane, K. W., & Linn, M. C. (2015). Automated, adaptive guidance for K–12 education. *Educational Research Review*. doi: 10.1016/j.edurev.2015.04.001.

- Gerard, L. F., Ryoo, K., McElhaney, K. W., Liu, O. L., Rafferty, A. N., & Linn, M. C. (2015). Automated Guidance for Student Inquiry. *Journal of Educational Psychology*, doi: 10.1037/edu0000052.
- Halloun, I., & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*, 53(11), 1043–1055.
- Hammer, D. (1996). Misconceptions or p-prims: How may alternative perspectives of cognitive structure influence instructional perceptions and intentions? *Journal of the Learning Sciences*, 5(2), 97–127.
- Herrenkohl, L. R., Tasker, T., & White, B. Y. (2011). Developing classroom cultures of inquiry and reflection using Web of Inquiry. *Cognition and Instruction*, 29(1), 1–44.
- Inhelder, B., & Piaget, J. (1969). *The early growth of logic in the child*. New York: Norton.
- Linn, M. C., & Eylon, B.-S. (2011). *Science learning and instruction: Taking advantage of technology to promote knowledge integration*. New York: Routledge.
- Linn, M. C., Gerard, L. F., Ryoo, K., McElhaney, K., Liu, L., & Rafferty, A. N. (2014). Computer-guided inquiry to improve science learning. *Science*, 344, 155–156. doi: 10.1126/science.1245980.
- Liu, O. L., Brew, C., Blackmore, J., Gerard, L. F., Madhok, J., & Linn, M. C. (2014). Automated scoring in inquiry science assessment: Application of c-rater. *Educational Measurement: Issues and Practice*. doi: 10.1111/emip.12028.
- McElhaney, K. W., Chang, H.-Y., Chiu, J. L., & Linn, M. C. (2015). Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education*, 51(1), 49–85. doi: 10.1080/03057267.2014.984506.
- National Academy of Education. (2009). *Education policy white paper on science and mathematics education*. J. Kilpatrick and H. Quinn (Eds.). Washington, DC: Author. Retrieved from <http://tinyurl.com/q593czo>.
- Nelson, H. (2013). *Testing more, teaching less*. Washington, DC: American Federation of Teachers. Retrieved from <http://tinyurl.com/qfqlb7n>.
- Richland, L. E., Linn, M. C., & Bjork, R. A. (2007). Instruction. In F. T. Durso (Ed.), *Handbook of applied cognition* (2nd ed.) (pp. 555–583). West Sussex, England: John Wiley & Sons, Ltd.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching*, 44(1), 57–84.
- Ryoo, K. L., & Linn, M. C. (2012). Can dynamic visualizations improve middle school students' understanding of energy in photosynthesis? *Journal of Research in Science Teaching*, 49(2), 218–243. doi: 10.1002/tea.21003.
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10, 176–199.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. doi: 10.1080/00461520.2011.611369.
- Vitale, J. M., Lai, K., & Linn, M. C. (2015). Taking advantage of automated assessment of student-constructed graphs in science. *Journal of Research in Science Teaching*. doi: 10.1002/tea.21241.
- Xie, Q., & Tinker, R. (2006). Molecular dynamics simulations of chemical reactions for use in education. *Journal of Chemical Education*, 83(1), 77–83.
- Zertuche, A., Gerard, L. F., & Linn, M. C. (2012). How do openers contribute to student learning? *International Electronic Journal of Elementary Education*, 5(1), 79–92. Retrieved from http://www.iejee.com/5_1_2012/IEJEE_5_1_79_92.pdf.
- Zhang, Z., & Linn, M. C. (2011). Can generating representations enhance learning with dynamic visualizations? *Journal of Research in Science Teaching*, 48(10), 1177–1198.

Rethinking and Redesigning Educational Assessment

*James W. Pellegrino*¹

Over the past 15+ years, members of the National Academy of Education (NAEd) have contributed to multiple reports and essays indicating the need to rethink basic assumptions underlying how we assess students, the role of assessment data in educational policymaking, and how we use assessment data to enhance teaching and learning. These reports and essays include Darling-Hammond et al. (2013); Glaser, Linn, and Bohrnstedt (1997); Gordon Commission (2013a,b); Kaestle (2013); Koretz (2009); Linn (2013); National Research Council (1999, 2001a, 2003, 2006, 2014); and Shepard (2000). These publications point to the critical importance of developing new kinds of classroom and large-scale assessments that work together to help all students learn and succeed in school by making as clear as possible to them, their teachers, and other education stakeholders the nature of their accomplishments and the progress of their learning.

I have been privileged to collaborate with many NAEd colleagues in developing the arguments and evidence presented in some of these reports. This essay draws on that background of work to consider why change is needed and how contemporary knowledge from behavioral and social science research, combined with advances in information technologies, can bring about the necessary and desired redesign of educational

¹ James W. Pellegrino is Liberal Arts and Sciences Distinguished Professor and Distinguished Professor of Education at the University of Illinois at Chicago. He was elected to the National Academy of Education in 2007.

assessment. An important caveat—I make no claim that my NAEd colleagues endorse any or all of what follows as articulated here.

Rising Expectations But Increasing Concerns About What and How We Assess

Expectations about what all students should learn—and, by implication, what they should be tested on—have changed significantly in response to social, economic, and technological changes. All students are now expected to demonstrate the kinds of reasoning and problem-solving abilities once expected of only a minority of young people. Assessments must therefore tap a broader range of competencies than in the past. They must capture the more complex skills and deeper content knowledge reflected in new expectations for learning. They must accurately measure higher levels of achievement while also providing meaningful information about students who still perform below expectations. These trends and pressures are being played out on a large scale and have accompanied the drive to set and meet challenging standards for student learning as reflected in the Common Core State Standards in Mathematics and English Language Arts and the Next Generation Science Standards (Achieve, 2013; Common Core State Standards Initiative, 2010a,b; National Research Council, 2012).

With the movement toward setting challenging academic standards and measuring students' progress in meeting those standards, assessment has been playing an increasing role in policy and decision-making. The amount of testing has increased significantly over the past decade given the advent of the No Child Left Behind legislation. In turn, many education stakeholders are now questioning whether current large-scale assessment policies and practices are yielding the most useful kinds of information for informing and improving education. At the same time, classroom assessments, which have the potential to enhance instruction and learning, are not being used to their fullest potential. Some of the typical frequently asked questions that now circulate in the media, research, and policy venues include: Are we testing too much? Are we testing the right things? What changes need to be made and how do we accomplish them? How should we envision the future of educational assessment?

Outmoded Theories and Underutilized Technologies

From teachers' informal quizzes to nationally and internationally administered standardized tests, assessments are an integral part of the educational process. In an ideal system, assessments help teachers, their students, and their students' parents determine how well students are

learning. They help teachers understand how to adapt instruction on the basis of evidence of student learning. They help principals and superintendents document the progress of individual students, classrooms, and schools. Finally, they help policymakers and the public make judgments about effectiveness of the investments that have been made in our educational systems.

Whether or not we realize it, every educational assessment, whether used in the classroom or large-scale policy context, is based on a set of scientific principles and philosophical assumptions. First, every assessment is grounded in a concept or theory about how people learn, what people know, and how knowledge and understanding progress over time. Second, each assessment embodies certain assumptions about which kinds of observations, or tasks, are most likely to elicit demonstrations of important knowledge and skills from students. Third, every assessment is premised on assumptions about how best to interpret the evidence from the observations in order to make meaningful inferences about what students know and can do.

Current assessment systems are the cumulative product of various prior theories of learning and models and methods of measurement. Although some of these foundations are still useful for certain functions of testing, considerable change is needed. The most common kinds of educational tests do a reasonable job with certain limited functions of testing, such as measuring knowledge of basic facts and procedures and producing overall estimates of proficiency for parts of the curriculum. However, both their strengths and limitations are a product of their adherence to theories of learning and measurement that are outmoded and fail to capture the breadth and richness of knowledge and competence. The limitations of these theories also compromise the usefulness of the assessments. Assessment systems should evolve to keep pace with developments in the sciences of learning and measurement if we are to achieve the learning goals embedded in many of our current standards.

Rethinking the Foundations of Assessment: The Merger of Cognition, Measurement, and Technology

Several decades of research in the learning sciences have advanced our knowledge about how children develop understanding in areas of the curriculum, how people reason and build structures of knowledge in academic subject areas, which thinking processes are associated with competent performance, and how knowledge is shaped by social context (representative work includes National Research Council, 1998, 2000, 2001b, 2004, 2007). For example, studies of expert-novice differences in subject domains have illuminated many critical features of proficiency that

should be the targets for assessment. Experts in a subject domain not only “know a lot,” but also, more importantly, they organize knowledge into schemas that support the rapid retrieval and application of such knowledge. Experts also use metacognitive strategies—ways of guiding one’s thinking—for monitoring understanding during problem-solving and for performing self-correction.

These and many other findings on how people learn and the differences in what novices and experts know suggest directions for revamping assessment practices to move beyond a focus on component skills and discrete bits of knowledge. Assessment should encompass the more complex aspects of student achievement. To aid learning, we need access to better information about students’ levels of understanding, their thinking strategies, and the nature of their misunderstandings.

During the past few decades significant developments have also accrued in measurement methods and theory. A wide array of statistical measurement methods are currently available to support the rigor we want in testing while simultaneously enabling the kinds of inferences about student knowledge that cognitive research suggests are important to pursue when assessing student achievement. In particular, it is now possible to characterize students in terms of multiple aspects of proficiency, rather than a single score; chart students’ progress over time, instead of simply measuring performance at a particular point in time; deal with multiple paths or alternative patterns of valued performance; model, monitor, and improve judgments based on informed evaluations; and report performance not only at the level of students, but also at the levels of groups, classes, schools, and states. Nonetheless, many of the newer models and methods are not widely used because they are not easily understood or made accessible for those without a strong technical background.

Technology offers the possibility of addressing this shortcoming. For instance, by building statistical models into technology-based learning environments for use in classrooms, teachers can assign more complex tasks, capture and replay students’ performances, share exemplars of competent performance, and in the process gain critical information about student competence. Without question, computer and information technologies are making it possible to create powerful learning environments and simultaneously assess what students are learning at very fine levels of detail, with vivid simulations of real-world situations, and in ways that are tightly integrated with instruction.

Research has already shown that assessments that inform teachers about the nature of student learning can help them provide better feedback to students, which in turn can significantly enhance learning. Many of the most effective examples of the use of assessment to inform learning

and instruction in the classroom rely on technology-based task presentation and information management systems.

If well-designed and used properly, assessments based on contemporary scientific knowledge could also promote more equitable opportunity to learn by providing better-quality information about the impact of educational interventions on children. More informative classroom assessments could result in earlier identification of learning problems and intervention for children at risk of failure, rather than waiting for results from large-scale assessments to signal problems. Students with disabilities could also benefit from this approach. At the same time, it is necessary for educators and researchers to continuously monitor the effects of their practices to ensure that the new assessments do not exacerbate existing inequalities.

Assessments based on contemporary theories and data on how competence develops across grade levels in a curriculum domain could also provide more valid measures of growth and the value added by teachers and schools. Such assessments could also enhance community dialogue about goals for student learning and important indicators of achievement at various grade levels and in different subject areas. Comparisons based on attainment of worthwhile learning goals, rather than normative descriptions of how students perform, could enhance the public's understanding of educational quality. New forms of assessment could also help provide descriptive and accurate information about the nature of achievement in a subject area and patterns of students' strengths and weaknesses that would be more useful than existing data for guiding policy decisions and reform efforts.

It is no surprise, then, that collective advances in the study of thinking and learning, in the field of measurement and in the deployment of powerful technologies for learning have stimulated many people to think in new ways about educational futures. New information technologies provide substantial opportunities to advance the design and use of assessments based on a merger of contemporary scientific knowledge of cognition and measurement. Focus is needed on ways to bring together the knowledge of how students learn, what they know and what is therefore worth assessing, with knowledge of how to do this with technical rigor, and ways to harness technology to make the merger feasible. Several intriguing implications arise from projecting what could happen from the coupling of advances in cognition, measurement and technology.

Visions of the Future

Extremely powerful information technologies are slowly becoming as ubiquitous in educational settings as they are in other aspects of people's

daily lives. They are almost certain to provoke fundamental changes in learning environments at all levels of the education system. However, many of the implications of technology are beyond people's speculative capacity—not that long ago few could have predicted the sweeping effects of the Internet on education and other segments of society. The range of computational devices and their applications is expanding exponentially, fundamentally changing how people think about communication, connectivity, information systems, educational practices, and the role of technology in society.

Although it is always risky to predict the future, it appears clear that advances in technology will continue to impact the world of education in powerful and provocative ways. Many technology-driven advances in the design of learning environments, which include the integration of assessment with instruction, will continue to emerge, and will reshape the terrain of what is both possible and desirable in education. Advances in curriculum, instruction, assessment, and technology are likely to continue to move educational practice toward a more individualized and mastery- or competency-oriented approach to learning. This evolution will occur across the K–16+ spectrum. To manage learning and instruction effectively, people will want and need to know considerably more about what has been mastered, at what level, and by whom.

Consider the possibilities that might arise if assessment is integrated into instruction in multiple curricular areas and the resultant information about student accomplishment and understanding is collected with the aid of technology. In such a world, programs of on-demand external assessment such as state achievement tests might not be necessary. Instead, it might be possible to extract the information needed for summative and program evaluation purposes from data about student performance continuously available both in and out of the school context.

Technology could offer ways of creating, over time, a complex data stream about how students think and reason while engaged in important learning activities. Information for assessment purposes could be extracted from this stream and used to serve both classroom and external assessment needs, including providing individual feedback to students for reflection about their learning strategies and habits. To realize this vision, additional research on the data representations and analysis methods best suited for different audiences and different assessment objectives would clearly be needed—and is certainly doable.

We can therefore imagine a future in which the audit function of assessments external to the classroom would be significantly reduced or even unnecessary because the information needed to assess students, at the levels of description appropriate for various monitoring purposes,

could be derived from the data streams generated by students in and out of their classrooms.

A metaphor for such a radical shift in how one “does the business of educational assessment” exists in the world of retail outlets, ranging from small businesses to supermarkets to department stores. No longer do these businesses have to close down once or twice a year to take inventory of their stock. Rather, with the advent of automated checkouts and barcodes for all items, these enterprises have access to a continuous stream of information that can be used to monitor inventory and the flow of items. Not only can business continue without interruption, but also the information obtained is far richer, enabling stores to monitor trends and aggregate the data into various kinds of summaries. Similarly, with new assessment technologies, schools would no longer have to interrupt the normal instructional process at various times during the year to administer external tests to students. Nor would they have to spend significant amounts of time preparing for specific external tests peripheral to the ongoing activities of teaching and learning.

Extensive technology-based systems that link curriculum, instruction, and assessment at the classroom level might enable a shift from today’s assessment systems, which use different kinds of assessments for different purposes, to a balanced design in which the three critical features of *comprehensiveness*, *coherence*, and *continuity* would be ensured. In such a design, assessments would provide a variety of evidence to support educational decision-making (*comprehensiveness*). The information provided at differing levels of responsibility and action would be linked back to the same underlying conceptual model of student learning (*coherence*) and would provide indications of student growth over time (*continuity*).

Clearly, technological advances will allow for the attainment of many of the goals that educators, researchers, policymakers, teachers, and parents have envisioned for assessment as a viable source of information for educational improvement. When powerful technology-based systems are implemented in classrooms, rich sources of information about student learning will be continuously available across wide segments of the curriculum and for individual learners over extended periods of time. This is exactly the kind of information we now lack, making it difficult to use assessment to truly support learning. The major issue is not whether this type of data collection and information analysis is feasible in the future. Rather, the issue is how the world of education anticipates and embraces this possibility, and how it explores the resulting options for effectively using assessment information to meet the multiple purposes served by current assessments and, most important, to enhance student learning.

It has been noted that the best way to predict the future is to invent it. Without doubt, multiple futures for educational assessment could be

invented on the basis of synergies that we know exist among information technologies and contemporary knowledge of cognition and measurement. While we are a considerable distance away from implementing the types of fully integrated systems envisioned above, there are steps that can be taken now that would put us on the path to a more productive future for educational assessment.

References

- Achieve. (2013). *Next generation science standards*. Retrieved from <http://www.nextgen-science.org>.
- Common Core State Standards Initiative. (2010a). *English language arts standards*. Washington, DC: National Governors Association and Council of Chief State School Officers. Retrieved from <http://www.corestandards.org/the-standards/english-language-artsstandards.pdf>.
- Common Core State Standards Initiative. (2010b). *Mathematics standards*. Washington, DC: National Governors Association and Council of Chief State School Officers. Retrieved from http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf.
- Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Aber, J. L., Baker, E., ... Steele, C. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education. Retrieved from <http://edpolicy.stanford.edu/publications/pubs/847>.
- Glaser, R., Linn, R., & Bohrnstedt, G. (1997). *Assessment in transition: Monitoring the nation's educational progress*. Washington, DC: National Academy of Education.
- Gordon Commission on the Future of Assessment in Education. (2013a). *Technical report*. Retrieved from http://www.gordoncommission.org/publications_reports.html.
- Gordon Commission on the Future of Assessment in Education. (2013b). *Policy report*. Retrieved from http://www.gordoncommission.org/publications_reports.html.
- Kaestle, C. (2013). *Testing policy in the United States: A historical perspective*. Princeton, NJ: Gordon Commission on the Future of Assessment in Education. Retrieved from http://www.gordoncommission.org/publications_reports/assessment_education.html.
- Koretz, D. (2009). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Linn, R. L. (2013). *Test-based accountability*. Princeton, NJ: Gordon Commission on the Future of Assessment in Education. Retrieved from http://www.gordoncommission.org/publications_reports/assessment_education.html.
- National Research Council. (1998). *Preventing reading difficulties in young children*. C. E. Snow, M. Burns, & M. Griffin (Eds.). Washington, DC: National Academy Press.
- National Research Council. (1999). *Grading the nation's report card: Evaluating NAEP and transforming the assessment of educational progress*. J. W. Pellegrino, L. Jones, & K. Mitchell (Eds.). Washington, DC: National Academy Press.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school* (expanded edition). J. D. Bransford, A. L. Brown, R. R. Cocking, M. S. Donovan, & J. W. Pellegrino (Eds.). Washington, DC: National Academy Press.
- National Research Council. (2001a). *Knowing what students know: The science and design of educational assessment*. J. W. Pellegrino, N. Chudowsky, & R. Glaser (Eds.). Washington, DC: National Academy Press.
- National Research Council. (2001b). *Adding it up: Helping children learn mathematics*. J. Kilpatrick, J. Swafford, & B. Findell (Eds.). Washington, DC: National Academy Press.

- National Research Council. (2003). *Assessment in support of learning and instruction: Bridging the gap between large-scale and classroom assessment*. Washington, DC: The National Academies Press.
- National Research Council. (2004). *Learning and instruction: A SERP research agenda*. M. S. Donovan & J. W. Pellegrino (Eds.). Washington, DC: The National Academies Press.
- National Research Council. (2006). *Systems for state science assessments*. M. R. Wilson & M. W. Bertenthal (Eds.). Washington, DC: The National Academies Press.
- National Research Council. (2007). *Taking science to school: Learning and teaching science in grade K–8*. R. A. Duschl, H. A. Schweingruber, & A. W. Shouse (Eds.). Washington, DC: The National Academies Press.
- National Research Council. (2012). *A framework for K–12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K–12 Science Education Standards, Board on Science Education. Washington, DC: The National Academies Press.
- National Research Council. (2014). *Developing assessments for the next generation science standards*. J. W. Pellegrino, M. Wilson, J. Koenig, & A. Beatty (Eds.). Washington, DC: The National Academies Press.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4–14.

Archaic Testing and Teaching in the United States: Why Do They Persist?

Robert J. Sternberg¹

Imagine if medical interventions today were comparable to those of a century ago. For example, until about 1910, mercury, a severe toxin, was used to treat syphilis. The disease, sometimes called “the great imitator because its symptoms mimicked the symptoms of so many other diseases,” was often hard to diagnose properly. Any physician today who used medical tests or treatments of a century ago certainly would be labeled a dangerous quack, or perhaps a medical Rip Van Winkle who just missed a century of progress.

In the field of psychometric testing, the tests, especially of intelligence, that are being given today are little different from the tests of a century ago. They certainly differ cosmetically, but they basically are cleaned up versions of very old tests. Large-scale educational interventions in the schools have not changed a whole lot either. Why have medical tests and treatments changed so much and psychometric tests and educational interventions so little?

First, there are many competing manufacturers of medical tests and of medications. These manufacturers need to innovate or die. The testing and textbook businesses, in contrast, have been monopolized by a few big players, sometimes, the same players for both businesses. Competition stimulates innovation but the competition in the testing field has been

¹ Robert J. Sternberg is Professor of Human Development at Cornell University. He was elected to the National Academy of Education in 2011.

minimal. What counts as an instructional innovation also amounts to little more than putting textbooks online.

Second, medications and medical tests are inevitably subject to a severe validation procedure—either they save lives or they do not; either they cure illness, or at least alleviate symptoms, or they do not. Standardized tests are evaluated by imperfect criteria (how valid, reliable, or even interesting are school grades or first-year college grade point averages [GPAs]?) via ambiguous statistics (just how good really is a correlation of 0.40?). And textbooks often end up being evaluated by performance on the standardized tests. So the evaluations of the highly imperfect interventions are by the even more imperfect tests.

Third, insurance companies will not pay for medical tests and procedures that are not subjected to rigorous tests leading to U.S. Food and Drug Administration (FDA) approval. In contrast, students and their parents pay for tests either through taxes or direct fees that they have little choice but to pay for. There is no FDA equivalent for tests or textbooks. Many colleges, for example, simply will not accept applications submitted without supporting scores on standardized tests. So students must pay or go elsewhere. Why the difference between medicine and education? The difference in stakes matters: No one dies as a result of poorly conceived textbooks or standardized tests. Someone's life may be ruined, but there is no unimpeachable statistic for a ruined life in the same way there is for a death. It is easier, therefore, to sell a third-rate educational test or textbook than a third-rate medical test or treatment.

Fourth, schools have been slow to adopt educational innovations. Even when there have been highly successful interventions (e.g., Palincsar & Brown, 1984), they have not been widely adopted. Whereas doctors are supposed to be constantly using innovative tests and treatments, the same expectation has not developed in our society with respect to schools. Quite the contrary: Innovations tend to be greeted with suspicion.

Finally, perhaps because medical research has attracted some of the best scientists, its researchers have actively strived for new, breakthrough tests and treatments, which have been put into practice. It has moved beyond the early 20th century. Strip away more sophisticated psychometric models and methods of scoring for psychometric tests, or many of the ebooks that simply put printed textbooks online, and one finds that education in practice has shown much less forward momentum, however far it may have come in theory. Today there is more use of computers, but too often students do on the computer merely what they used to do with pencil and paper.

The content of psychometric tests and educational texts changes, but the changes often are more market-driven than science-driven. For example, writing tests have come on and gone off the SAT as a function of

political pressures (e.g., eagerness not to lose a contract with the University of California system). And state boards that decide on what constitute acceptable textbooks are often driven more by political or even religious considerations than they are driven by scientifically sound content or pedagogy. In 2015, these boards are still deciding whether evolution is anything more than just a speculative scientific theory, based on religious and political considerations.

With regard to standardized tests, in particular, there are a variety of societal pressures that have tended to keep conventional standardized tests enthroned. First, the tests give the appearance of quantitative precision, always a valuable commodity when a company wishes a test to appear scientific. Second, the people making admissions or placement decisions about test-takers, based on the test-takers' results, generally did well enough on these tests to get themselves into positions where they can make these judgments. They end up looking for people like themselves, that is, people who can succeed on standardized tests. Third, publication of various ratings and rankings that heavily weigh test scores leads schools, colleges, and universities to place more and more emphasis on test scores in their assessment processes, raising the importance of tests to the institutions and society. Fourth, use of tests is entrenched, and it simply is difficult to get people to change what they already are doing unless there is a compelling reason to do so. Fifth, when tests are used to make high-stakes decisions, only people who test well will get into coveted positions, creating self-fulfilling prophecies. The powerful people in society often did well on tests because those with lower test scores who might otherwise have succeeded were blocked from advancement, creating the dubious impression that elevated test scores are needed for an individual to be successful. Finally, people in positions of authority vouch for the tests, an imprimatur that creates belief in the tests by dint of the educators' appearance of expertise and authority.

Can we do better in our assessment and instruction? I believe we can, using assessment and instruction based on a theory of successful intelligence (Sternberg, 1997, 2003). The basic idea underlying this theory is that whereas traditional tests of abilities and achievement assess almost exclusively memory and analytical abilities, a more nearly complete test would need to assess creative, practical, and even wisdom-based abilities as well. Similarly, a school curriculum that taught young people how to succeed in the everyday world also would have to develop the diverse thinking skills underlying these (modifiable) abilities.

Consider, as an example, college admission testing (in analogy, at the beginning of the article, to medical testing). In three successive projects, *Rainbow*, *Kaleidoscope*, and *Panorama*, my colleagues and I have shown that it is possible to separate creative and practical ability factors from an

analytical (g-based) one (increased construct validity), increase prediction of college GPA over that attained by standardized test scores and high school grades (increased predictive/concurrent validity), substantially decrease effects of ethnic-group differences on test scores (increased equity), and increase acceptability to examinees and their parents with respect to the holistic potential of the applicants for college (increased face validity) (see Sternberg, 2010, 2015; Sternberg, Bonney, Gabora, & Merrifield, 2012). My colleagues and I also have shown that it is possible to expand achievement tests used in college admissions to include assessments of creative and practical as well as memory and analytical abilities, thereby reducing ethnic-group differences (Stemler, Grigorenko, Jarvin, & Sternberg, 2006; Stemler, Sternberg, Grigorenko, Jarvin, & Sharpes, 2009).

Consider further instruction in schools (in analogy, at the beginning of the article, to medical interventions). On the one hand, my colleagues and I have shown that teaching for successful intelligence (as described by Sternberg & Grigorenko, 2007) can increase school achievement (Sternberg, Torff, & Grigorenko, 2008), especially if the teaching is matched at least some of the time to correspond to students' cognitive strengths (Sternberg, Grigorenko, Ferrari, & Clinkenbeard, 1999). On the other hand, we have found that massive scaling-up efforts require greater attention to teacher training and intervention fidelity than we so far have achieved (Sternberg et al., 2014).

When I started out my career, I was hopeful that I would see during my lifetime significant changes across the country in assessment and instruction, perhaps partly as a result of my efforts (Sternberg, 2014, in press) or the efforts of others such as Howard Gardner (e.g., Gardner, 1983/2011). That has not happened. Although there have been sporadic improvements, the No Child Left Behind Act has had extremely damaging effects on educational progress, supported as it has been by mostly well-meaning politicians and clueless educators whose vision is locked into the early 20th century. But I do not believe this state of affairs is a cause for despair.

My undergraduate advisor, Endel Tulving, once told me that it is extremely difficult to get anything in the world to change, and so I hope that my students or the students of other professors succeed in effecting the changes that my generation so far has been unable to bring to fruition.

References

- Gardner, H. (1983/2011). *Frames of mind: The theory of multiple intelligences*. New York: Basic.
- Palincsar, A. S., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, 1(2), 117–175.

- Stemler, S. E., Grigorenko, E. L., Jarvin, L., & Sternberg, R. J. (2006). Using the theory of successful intelligence as a basis for augmenting AP exams in psychology and statistics. *Contemporary Educational Psychology, 31*(2), 344–376.
- Stemler, S., Sternberg, R. J., Grigorenko, E. L., Jarvin, L., & Sharpes, D. K. (2009). Using the theory of successful intelligence as a framework for developing assessments in AP Physics. *Contemporary Educational Psychology, 34*, 195–209.
- Sternberg, R. J. (1997). *Successful intelligence*. New York: Plume.
- Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. New York: Cambridge University Press.
- Sternberg, R. J. (2010). *College admissions for the 21st century*. Cambridge, MA: Harvard University Press.
- Sternberg, R. J. (2014). I study what I stink at: Lessons learned from a career in psychology. *Annual Review of Psychology, 65*, 1–16.
- Sternberg, R. J. (2015). The Rainbow Project and beyond: Using a psychological theory of intelligence to improve the college admissions process. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, & J. R. Pomerantz (Eds.), *Psychology and the real world* (2nd ed., pp. 139–146). New York: Worth.
- Sternberg, R. J. (in press). Still searching for the Zipperumpazoo. *Child Development Perspectives*.
- Sternberg, R. J., & Grigorenko, E. L. (2007). *Teaching for successful intelligence* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- Sternberg, R. J., Torff, B., & Grigorenko, E. L. (1998). Teaching triarchically improves school achievement. *Journal of Educational Psychology, 90*, 374–384.
- Sternberg, R. J., Grigorenko, E. L., Ferrari, M., & Clinkenbeard, P. (1999). A triarchic analysis of an aptitude–treatment interaction. *European Journal of Psychological Assessment, 15*(1), 1–11.
- Sternberg, R. J., Bonney, C. R., Gabora, L., & Merrifield, M. (2012). WICS: A model for college and university admissions. *Educational Psychologist, 47*(1), 30–41.
- Sternberg, R. J., Jarvin, L., Birney, D., Naples, A., Stemler, S., Newman, T., Otterbach, R., Randi, J., & Grigorenko, E. L. (2014). Testing the theory of successful intelligence in teaching grade 4 language arts, mathematics, and science. *Journal of Educational Psychology, 106*, 881–899.

Challenges in Higher Education

Women in the Academy: Past, Present, and Future

*Stephen J. Ceci*¹

Of the challenges facing postsecondary education policymakers, the ones that get the most attention are economic—tuition increases that outpace inflation and unsustainable wages paid to contingent faculty. In this essay I briefly describe what I expect will be the largest challenge for the next generation of higher education policy scholars, and it has little to do with economics.

I begin with some actuarial data from a National Research Council (NRC) report (National Research Council, 2009). Its findings accord with those from other large-scale actuarial analyses, several dating back to the 1980s. Table 1 shows that in all six fields analyzed by the NRC panel, only 11 percent to 26 percent of applicants for professorships were women. Yet those women were offered professorships at a higher rate than men.

My colleagues and I have reviewed many large-scale actuarial analyses; like that of the NRC, all of them reveal a hiring advantage for female applicants for tenure-track professorships: women are a smaller fraction of the applicant pool, but those who apply are more likely to be hired. This hiring preference for women goes beyond the six fields analyzed in the NRC study, including other fields in which women are underrepresented—computer science, economics, and geosciences.

Colleagues, journal reviewers, and policymakers are usually surprised by these findings because they conflict with the sexist hiring nar-

¹ Stephen J. Ceci is the Helen L. Carr Professor of Development Psychology at Cornell University. He was elected to the National Academy of Education in 2015.

TABLE 1 Fraction of Female Applicants for Tenure Track Positions Invited to Interview and Offered Positions at 89 U.S. Research Universities

Field	Mean % Female Applicants	Mean % Invited to Interview	Mean % Offered Position
Physics	12%	19%	20%
Biology	26%	28%	34%
Chemistry	18%	25%	29%
Civil Engineering	16%	30%	32%
Electrical Engineering	11%	19%	32%
Mathematics	20%	28%	32%

SOURCE: National Research Council, 2009, p. 8, Findings 3-10, 3-13.

rative made salient in studies that show women are discriminated against in hiring, mentoring, and remuneration of lab managers (Moss-Racusin, Dovidio, Brescoll, Graham, & Handelsman, 2012), the awarding of post-doctoral fellowships (Wenneras & Wold, 1997), and hiring of entry-level psychologists (Steinpreiss, Anders, & Ritzke, 1999). My colleagues and I have spent considerable time trying to reconcile these findings with those of the actuarial reports from the NRC and others that show women are preferentially hired.

To some, this pro-female hiring advantage is not due to a preference for women among faculty search committees. Rather, it is due to female applicants for tenure-track posts being stronger, on average, than their male competitors. The argument is that they survived the sexist winnowing process in college and graduate school, with fewer role models and mentors and access to psychological resources. Such claims are omnipresent. Consider the following:

It is important to note that (women’s) higher rates of success do not imply favoritism, but may be explained by the possibility that only the strongest female candidates applied for Research I positions. This self-selection by female candidates would be consistent with the lower rates of application by women to these positions. (National Research Council, 2009, p. 54)

Perhaps the women who survive training in a field where they have few mentors and surmount barriers most men may have little knowledge of, might actually be better. At least we cannot assume they aren’t. (Jacobson, 2011)

My colleagues and I have attempted to shed light on this issue and the claims surrounding it. One of our studies was a monograph-length jour-

nal article (>47,000 words), reporting hundreds of analyses (Ceci, Ginther, Kahn, & Williams, 2014). We examined gender differences in domains such as salary, tenure and promotion rates, persistence (in years on the job), productivity, citations, job satisfaction. We used national databases such as the National Science Foundation Survey of Doctorate Recipients. Additionally, we disaggregated the data by field, type of institution, and rank, so for each analysis there were 8–18 fields for each of three ranks in research-intensive versus teaching-intensive institutions. These analyses provided a finer-grained picture, revealing many field-specific findings:

To foreshadow our conclusions, we find that although women are under-represented in GEEMP (geosciences, engineering, economics, mathematics, physical sciences) fields, the overall state of the academy (collapsing across the many hundreds of between-sex contrasts involving salary, promotion, type of institution, type of field, and transition points) is largely one of gender neutrality—with some notable exceptions that should be of interest to members of specific fields. (Ceci et al., 2014, p. 77)

We found, for example, that tenure rates were typically the same for women and men, with the exception of the field of biology, where men were tenured more often than women. In 18 out of the 24 rank-field comparisons among faculty at research-intensive universities (R1s), there were no gender differences in salary (Ceci et al., 2014, Table 4, p. 117), and persistence in terms of years on the job was comparable for both genders.

We did not examine the hiring domain. The actuarial findings from the NRC and others are open to competing interpretations, such as women are hired more often because they are stronger applicants versus women are preferred over identically strong men because faculty value gender diversity. Resolving these competing interpretations requires an experimental approach. In five national experiments (Williams & Ceci, 2015), we randomly sampled 872 tenure-track faculty from 371 institutions in all 50 states. Faculty were asked to make hiring choices, usually between two male applicants and one female applicant. The applicants were presented in such a way to mask the experiments' goal: faculty were led to believe that we were interested in their preference between identically qualified applicants who differed in personality such as creative versus analytic applicants. However, these personality traits were counterbalanced between the applicants, along with counterbalanced lifestyle information. The ruse worked, and few faculty were aware that the experiments were about pro-female hiring preference.

The results of the five experiments were fascinating; all sorts of interesting findings emerged, such as male faculty preferred to hire a married father of two young children who had a stay-at-home partner over an identically qualified mother of two children who was divorced. Female

faculty preferred to hire women whose records showed no family leave during graduate school, whereas male faculty preferred women who took family leaves. (Neither gender of faculty cared whether male applicants took family leaves.) Everyone preferred to hire a single woman without children over a single man without children.

Across the myriad experimental conditions, faculty of both genders strongly preferred to hire women over identically qualified men at a ratio of 2 to 1. Coupled with the actuarial data showing that female applicants are more likely to be hired in the real world, these experimental data suggest it is a propitious time for talented women to launch tenure-track careers—not because they are stronger applicants (they were identical in our experiments), but because faculty appear to have internalized the value of gender diversity (e.g., having female role models for female majors). Of course, this pro-female hiring preference says nothing about barriers that may occur post hiring. In our monograph-length analysis we concluded that, for the most part, the playing field is gender-neutral in terms of salary, tenure, promotion, job satisfaction, and persistence. It was not always this way; in the not-too-distant past the playing field favored men in all domains and women were shut out of many professions. However, those days are over, and today even the most male-dominated fields such as computer science, physics, and engineering eagerly pursue women applicants. For example, the Computing Research Association reported the results of a national study of hiring:

As new Ph.D.s, women submitted far fewer applications than men and received many more offers per application. Female new hires applied for only 6 positions (compared with 25 for men), obtained 0.77 interviews per application (vs. 0.37 for men), and received 0.55 offers per application (vs. 0.19 for men). Obviously, women were much more selective in where they applied, and also much more successful in the application process. (Stankovic & Aspray, 2003, p. 31)

Of course, challenges related to family formation continue to affect women far more than men. In the future the National Academy of Education will need to shift from its past emphasis on hiring hurdles that young pre-tenure mothers face. Mandating that a certain fraction of search committees be composed of women or that all search members undergo gender awareness training, as many universities do, does not address the real issues facing female faculty today. Gender sensitivity training may have made sense in the past, but in recent decades the hiring process has been favorable to women applicants, not only in experiments but also in real-world academic hiring. Salaries and startups are typically gender-fair, as are teaching loads and promotions. Colleagues who have carved niches at their institutions based on such initiatives may be reluctant to abandon

them, but I predict that the findings described here will become increasingly widely known and administrators will at some point shift from interventions focused on the point of hiring. The challenge is to replace them with initiatives that address family-related hurdles that women face. Current initiatives such as stopping the tenure clock for family-related issues may prove unsatisfactory. For example, recent research shows that stopping the tenure clock does not increase women's probability of being promoted, but it does reduce salaries of women who stopped the clock (Manchester, Leslie, & Kramer, 2013).

References

- Ceci, S. J., Ginther, D. K., Kahn, S., & Williams, W. M. (2014). Women in academic science: A changing landscape. *Psychological Science in the Public Interest*, 17, 76–146. doi: 10.1177/1529100614541236.
- Jacobson, A. J. (2011, February 17). Is bias “yesterday’s problem”? Addition. Retrieved from <https://feministphilosophers.wordpress.com/2011/02/17/is-bias-yesterdays-problem>.
- Manchester, C.F., Leslie, L.M., & Kramer, A. (2013). Is the clock still ticking? An evaluation of the consequences of stopping the tenure clock. *Industrial and Labor Relations Review*, 66(1).
- Moss-Racusin, C., Dovidio, J., Brescoll, V., Graham, M., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 16474–16479.
- National Research Council. (2009). *Gender differences at critical transitions in the careers of science, engineering, and mathematics faculty*. Washington, DC: The National Academies Press.
- Stankovic, J., & Aspray, W. (2003). Recruitment and retention of faculty in computer science and engineering. Computing Research Association. Retrieved from <http://archive.cra.org/reports/r&rfaculty.pdf>.
- Steinpreis, R., Anders, R. K., & Ritzke, K. D. (1999). The impact of gender on the review of the CVs of job applicants and tenure candidates: A national empirical study. *Sex Roles*, 41, 509–528.
- Wenneras, C., & Wold, A. (1997). Nepotism and sexism in peer-review. *Nature*, 387, 341–343.
- Williams, W. M., & Ceci, S. J. (2015). National hiring experiments reveals 2-to-1 preference for women faculty on STEM tenure-track. *Proceedings of the National Academy of Sciences of the United States of America*. Retrieved from <http://www.pnas.org/content/early/2015/04/08/1418878112.abstract>.

Learning What Research Says About Teacher Preparation

*Robert E. Floden*¹

A perennial question for education leaders and policymakers is how initial teacher preparation should be arranged so that those completing preparation are of high quality and are likely to be successful in teaching a wide range of students. In the last half of the 20th century, scholars, practitioners, and social commentators wrote volumes characterizing the state of teacher education in the United States. Some insisted that most teacher preparation programs were doing a poor job, and made sweeping claims about what should change (e.g., Conant, 1963). Others were somewhat more hopeful about the preparation that many teachers were receiving, but nonetheless made strong recommendations for changes (e.g., Goodlad, 1990). Some of what was written drew on evidence, but the evidence was always drawn from a sample of programs, and relied more on opinions about what program features were most desirable than on systematic research designed to accurately describe the national picture or to estimate the effects of program characteristics on program graduates' knowledge or subsequent performance. The volumes written often received substantial attention, sometimes achieving visibility among policymakers or even among a wider public, at other times being discussed mostly within a smaller circle of those with particular interest in teacher education.

¹ Robert E. Floden is University Distinguished Professor and Associate Dean for Research, College of Education at Michigan State University. He was elected to the National Academy of Education in 2007.

Starting roughly at the turn of the millennium, a series of projects were undertaken to identify trustworthy relevant research and what conclusions about teacher preparation that research would support. Most of the projects were staffed in part by scholars who were at the time members of the National Academy of Education (NAEd), or who, like me, would later be elected to the Academy.

These efforts were notable both for the care taken to identify and summarize the available research and for their frequent conclusion that the available body of research was “thin.” I will sketch several of these projects, then suggest why the body of research has grown slowly and why I see hope for speedier future growth.

Early in 2000, the U.S. Department of Education asked the Center for the Study of Teaching and Policy at the University of Washington to carry out a study that would give the Department a better understanding of what claims about teacher education were supported by research, and what claims were not. Department staff were hearing many claims, often ostensibly supported by research, but were doubtful that most claims had solid empirical support. The Center in turn, asked my colleagues and me to carry out a careful literature review. The questions for which they wished to have answers included: What kinds of subject-matter preparation, and how much of it, do prospective teachers need? What are the components and characteristics of high-quality alternative certification programs? What kinds, timing, and amount of clinical training (“student teaching”) best equip prospective teachers for classroom practice? Because the Department wished to have answers to these questions before the end of the Presidential administration (i.e., January 2001), the work had to be completed on a short timeline—4 months.

To ensure that our review was viewed by policymakers as truly representative of research, rather than dictated by some political agenda, the Department worked with us to put together a technical working group of distinguished scholars to provide advice about many aspects of our review, from criteria for what research to include to whether our reports were phrased in ways that reflected the available evidence. The advisory group, all of whom are now members of the NAEd, included Bruce Alberts and Ellen Condliffe Lagemann, at the time presidents of the National Academy of Sciences and the Spencer Foundation, respectively. Our team examined more than 300 published reports and identified 57 that met the criteria we had established for inclusion. Our report (Wilson, Floden, & Ferrini-Mundy, 2001), which has been cited more than 800 times, found that “overall, the research base concerning teacher preparation is relatively thin.”

At the same time we were working on this review, the American Educational Research Association created a Panel on Research and Teacher

Education, with a similar charge, but longer time frame—a period of years. The Panel was chaired by Marilyn Cochran-Smith and Ken Zeichner (2005), both leading scholars who were subsequently elected to the NAEd. Once again, the intent was to summarize what was known from research (“the weight of evidence”), as a way of informing the lively debates about teacher preparation, which were characterized by multiple competing claims. This larger project considered results from a larger number of studies, addressing a broader set of questions, but once again focused on what could be learned from rigorous research about the national picture of teacher preparation and the effects of teacher preparation program characteristics on the knowledge, skill, and performance of program completers. The Panel identified some studies that connected program characteristics to program outcomes, such as that when teachers supervising student teachers and university teacher educators are congruent in the teaching practices they support (as opposed to giving conflicting advice), student teachers are more likely to successfully implement those practices. But a frequent conclusion drawn in the chapters of this volume was that relatively little trustworthy research was found connecting features of teacher preparation to changes in teacher preparation students, particularly changes in practice.

At roughly the same time, the NAEd had a project aimed at summarizing research on topics important to the curriculum of teacher preparation. The volume produced (Darling-Hammond & Bransford, 2005) devoted chapters to topics including theories of learning, assessment, and how teachers learn. It differed from the projects mentioned above in that it summarized research on topics seen to be important components of the curriculum of teacher preparation, rather than looking for research on what effect studying these topics had on teachers, teaching, and ultimately student learning. That is, it was a compendium of research on topics seen as important for teachers to know about, rather than research on the effects of including these topics in the teacher preparation curriculum on what program graduates did in their classrooms. It thus was able to draw on several large bodies of research, but what it said about the national state of teacher preparation, or about the effects of teacher preparation programs on important outcomes, was mostly restricted to illustrative descriptions of a sample of programs.

A final report on research on teacher preparation was commissioned by Congress and carried out as a consensus panel study by the National Research Council (National Research Council, 2010). Congress specified the questions to be addressed, including: What sorts of instruction and experiences do teacher candidates receive in preparation programs of various types? In addition, to what extent are the required instruction and experiences consistent with converging scientific evidence? The panel

members were distinguished scholars, including several NAEd members. Once again, an overall conclusion of the report is that research on teacher preparation is comparatively thin.

To sum up, in the past 15 years, several initiatives have been undertaken to carefully examine the research evidence about initial teacher preparation in the United States, with particular attention to assessing what is known about the connections between features of teacher preparation programs and outcomes of those programs. Those initiatives, in which the NAEd members have played important roles, represent an important change from prior publications about teacher preparation, which were based on some combination of the authors' personal views and descriptions of small samples of programs. These new initiatives were able to identify relevant high-quality studies, but, in comparison to the bodies of evidence about other important topics in education, the bodies of evidence about teacher preparation were thin. We now have a much clearer picture of what is known about the national state of teacher preparation and about the effects of various approaches used in such programs. What is clear is that we know much less than the strong claims by critics and supporters would suggest.

What should we make of this? Why is evidence about teacher preparation thin? Is there hope that the body of research is getting thicker?

I suggest three main reasons that the body of evidence is comparatively thin. First, the institutional organization of research in the United States has been slow to encourage research on teacher education. Schools of education have often separated staff working in teacher preparation from staff given the most time and encouragement to carry out research. Federal funders have sometimes believed that studies of teacher education were not a good investment. Therefore, to some extent, the comparatively small number of solid studies describing U.S. teacher preparation as a whole or documenting effects of teacher preparation practices is due to a relatively small investment in research on these topics. You get what you pay for.

Second, the institutional context of teacher preparation makes research difficult. Most initial teacher preparation takes place within higher education, where it is generally more difficult to study connections between instruction and outcomes than it is in elementary and secondary school. One problem is that astounding variation in higher education curriculum, organizational structure, institutional mission, and student population. That variation may be even greater in teacher preparation, where dozens of distinct preparation programs often operate within a single university. Another problem is that adult students learn from a wider variety of sources than do young school children. Teacher education students learn from their parents, partners, former teachers, in addition to what goes

on in formal program experiences. In addition, as Lortie (1975) stressed, teachers enter formal programs having already learned much from their “apprenticeship of observation,” the dozen years in elementary and secondary school when they were closely observing their own teachers, drawing conclusions on which they rely for their own teaching practice.

Third, the absence of widely accepted measures of program characteristics and outcomes has made it difficult to assemble data sets useful for research. With few exceptions, the curriculum of teacher preparation programs is idiosyncratic to individual institutions and often to individual instructors, with uniformity only at the level required to meet state and national program approval standards. States bring some commonality to outcome measures through certification examinations, but the grain size of those examinations limits research.

Some recent developments, however, give grounds for hope that the depth of research is increasing. The recent TEDS-M study of the preparation of mathematics teachers (Blömeke, Hsieh, Kaiser, & Schmidt, 2014; Tatto et al., 2012) has demonstrated the possibility for gathering nationally representative data on teacher preparation programs and outcomes. It also demonstrated that federal agencies are willing to invest in research on teacher preparation. In the recent Institute of Education Sciences (IES) research competition, the call for proposals explicitly included research on preservice teacher preparation as a topic of interest. As a followup to the 2010 report, the National Research Council has begun to explore possibilities for work that could answer the calls for new research in that report. Members elected to the NAEed in the past decade have included several scholars whose research focuses on teacher preparation. Data systems in several states now include information on students’ experiences in teacher preparation and on their subsequent classroom performance. All of these developments portend well for addressing the issues that have slowed past progress. Large-scale investments have been made. Scholarship about teacher preparation is being nationally recognized. Greater attention is being given to research on postsecondary education in general, which will likely support additional work on teacher preparation. Progress is also evident in building data systems that permit cross-institutional research.

Interest in questions about the effects of varying approaches to teacher preparation is high. If a set of research summaries is commissioned in a decade, then perhaps the conclusion about the depth of research will be more positive.

References

- Blömeke, S., Hsieh, F.-J., Kaiser, G., & Schmidt, W. H. (Eds.). (2014). *International perspectives on teacher knowledge, beliefs and opportunities to learn*. New York: Springer.
- Conant, J. B. (1963). *The education of American teachers*. New York: McGraw-Hill.
- Darling-Hammond, L., & Bransford, J. D. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Jossey-Bass.
- Goodlad, J. I. (1990). *Teachers for our nation's schools*. San Francisco, CA: Jossey-Bass.
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. Chicago, IL: University of Chicago Press.
- National Research Council. (2010). *Preparing teachers: Building evidence for sound policy*. Committee on the Study of Teacher Preparation Programs in the United States, Center for Education, Division of Behavioral and Social Sciences. Washington, DC: The National Academies Press.
- Tatto, M. T., Schwille, J., Senk, S. L., Ingvarson, L., Rowley, G., Peck, R., ... Reckase, M. (2012). *Policy, practice, and readiness to teach primary and secondary mathematics in 17 countries: Findings from the IEA Teacher Education and Development Study in Mathematics (TEDS-M)*. Amsterdam, Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2001). *Teacher preparation research: Current knowledge, gaps, and recommendations*. Seattle, WA: Center for the Study of Teaching and Policy.
- Zeichner, K. M., & Cochran-Smith, M. (2005). *Studying teacher education: The report of AERA Panel on Research and Teacher Education*. Mahwah, NJ: Lawrence Erlbaum.

Federal Involvement in Higher Education Finance

*Michael McPherson*¹

The Higher Education Act (HEA), adopted 50 years ago as part of Lyndon Johnson's Great Society initiative, had only a modest immediate impact, but its symbolic import was immense. Apart from the two very important Morrill Acts of 1867 and 1890, federal engagement in higher education had until then been limited to programs with a close military connection, including the World War II and Korean War GI Bills, and the National Defense Education Act, which employed a Cold War rationale to create National Defense Student Loans (NDSLs, now Perkins loans).

With the HEA, Lyndon Johnson established federal investment in higher education as a matter of national importance in its own right, notably in connection with expanding educational opportunity. At the Act's signing, he declared, extravagantly, "[This Act] means that a high school senior anywhere in this great land of ours can apply to any college or any university in any of the 50 States and not be turned away because his family is poor."² Two main elements of the Act addressed the expansion of college opportunity. First, and perhaps most consequentially, the HEA created a system by which the federal government would encourage banks to make loans available to needy students on favorable terms by guaranteeing the banks that the federal government would make up any

¹ Michael McPherson is the President of the Spencer Foundation. He was elected to the National Academy of Education in 2005.

² President Johnson made these remarks on November 8, 1965, at Southwest Texas State College upon signing the Higher Education Act of 1965.

losses they suffered due to non-payment. Second, the federal government created a system of awards to colleges with the funds earmarked to help needy students, and requiring matching funds from the colleges. These awards included work-study job opportunities on campus and so-called Educational Opportunity Grants (EOGs), to be awarded by the colleges to exceptionally needy students. The HEA also incorporated Eisenhower's earlier NDSLs into its package of awards, which over time came to be called the "Campus-Based" Programs.

The great significance of this legislation was that it "broke the ice" by establishing the principle that the federal government had a continuing peace-time role in higher education finance. Of particular note is the fact that the loan program took the legislative form of an entitlement; the government committed itself to provide sufficient funds for the program until such time as the Congress changed the law. In its early incarnation the program was limited to students with demonstrated financial need and included a subsidy in the form of waiver of interest payments while students were in school. Entitlement programs, of which Medicare and Social Security are prime examples, have proved likely to grow in cost and participation over time, and government loans—now provided directly instead of through banks and not limited to students with financial need—are no exception.

EOGs and the other Campus-Based Programs had the distinctive feature that, although the spending was designed to benefit students, it was actually provided to colleges and universities themselves, which were then obliged to distribute the funds according to rules that allowed the colleges some discretion. This hybrid arrangement was a compromise between the preferred approaches of two different groups. One group of actors, closely associated with private colleges and universities, and with the College Board, preferred that money go directly to students, who could then "vote with their feet" about where to use the funds. The alternative view, espoused by the American Council on Education among others, was that money should be awarded directly to deserving colleges, which would have wide discretion about how to spend it. The 1965 compromise between these views was likely key to getting the general principle of sustained federal support of higher education into law.

That compromise, though, did not stick. When the time came to renew the HEA in 1972, Congress wrestled mightily over the question of institutional aid versus student aid as the principal federal vehicle for supporting higher educational opportunity.³ Supporters of a strong insti-

³ Larry Gladieux and Tom Wolanin provided a wonderfully informative and interesting account of this entire episode in their book, *Congress and the Colleges: National Policies of Higher Education* (part of the Lexington Books Politics of Education Series in 1976).

tutional role rallied around Edith Green, an Oregon member of the U.S. House of Representatives, while the supporters of aid directly to students were led by Claiborne Pell of Rhode Island. Both were Democrats. With support from the Nixon administration falling on the side of awards to students, Congress in the Education Amendments of 1972 introduced a new "Basic Educational Opportunity Grant" (BEOG) program, with awards going directly to students, based on their family's ability to pay and the cost of their education. In classic American fashion, the earlier Campus-Based Programs were preserved, with the old EOGs renamed Supplemental Educational Opportunity Grants. BEOGs of course later became Pell grants.

Remarkably, in the course of 7 years, the U.S. government had gone from having no continuing role in the finance of undergraduate higher education to inventing an entire new structure of federal financing whose main elements remain in place today. Viewing these remarkable years from the standpoint of the National Academy of Education, with our interest in education policy and scholarship, what might we notice?

First, the widely expressed discontent about higher educational policy these days should be viewed not as a sign of failure but as a by-product of the remarkable success of these policies. In 1965 a low-income family without assets simply could not borrow at a reasonable interest rate for education, and its children often had no realistic chance of beginning college. Thus the introduction of loans unambiguously expanded opportunity. Opportunity is now so widespread that it is increasingly thought that students from families of limited means should be able to attend college without needing to borrow. That this should be a plausible position even to consider is testimony to the amount of progress that has been achieved.

Even as access to capital has grown, the Pell grant program has continued to expand over the decades, in both Democratic and Republican administrations, both in the number of students served and the average size of grants. Meanwhile state governments over the past two decades have substantially reduced the per-student funding they provide to public institutions. In a major reversal of roles, the federal government is now a larger source of funding for higher education than are the states. Not surprisingly, this great expansion in the federal role has led to a growing public expectation that the federal government should be held accountable for producing good results on such metrics as college prices, graduation rates, and debt levels.

In this way, the striking success of federal student aid programs has put the government in a tricky spot. In 1972, most of the public financing of higher education came from states' appropriations to public colleges and universities, institutions that competed with one another and with private colleges. Congress could have inserted itself into that competi-

tion by sending money to colleges in proportion to criteria established by federal policymakers. Instead, Congress elected to be “impartial”⁴ about the merits of institutions and put the money into the hands of low-income students, enabling them, like their more affluent counterparts, to decide for themselves where to attend. Had Congress opted for institution-based aid, they might, for example, have provided more federal support to institutions that kept their tuitions down, or had a particularly large representation of low-income students, or a high graduation rate. Under the Pell grant voucher scheme, it would be the collective decision of students voting with their feet that would determine how influential prices or graduation rates would be in determining institutions’ success. Opting for the Pell program amounted to a decision to give the market (and the states, through their operation of public institutions) a larger role and explicit federal administrative decisions a smaller role in determining the policies institutions set and the relative success of different higher education sectors.

That more or less “hands off” approach worked pretty well when federal funding played a relatively small role in higher education finance, but now that the federal government is the major player, there is a greater expectation that Congress and the Administration will take an active role not only in funding students but also in producing results. This is a hugely difficult challenge. The federal government obviously does not operate colleges and universities (excepting the service academies) and in fact, given the path chosen in the 1970s, does not even fund institutions directly. Its tools for modifying institutions’ behavior are by design limited. Its efforts to overcome those limits often prove clumsy—as illustrated by the ill-fated effort to create a federal rating system for colleges.

The challenge of creating a more accountable higher education system may well be the next frontier for higher education finance, and that challenge could result in significant modification of the existing voucher approach in federal policy, especially if the state role continues to shrink and the federal role to expand. But even if substantial change may be in the offing, it is important to recognize the substantive achievements of federal higher education policy in the student aid era. Perhaps even more important is recognizing how much these policies have helped engrain in the American consciousness the presupposition that ability to pay should

⁴ I recognize that “impartial” is a loaded word here—different policies distribute benefits to institutions and to students differently. The Pell grant program as it has evolved is “impartial” in the sense that the amount of support provided to a student depends principally on characteristics of the student and her family, with the characteristics of the institution attended playing only a very limited role.

not be an insurmountable barrier to access or, increasingly, to success in college for qualified students.

I will conclude by noting some important contributions of social science research in shaping both the HEA and the Education Amendments. Economists such as Ted Schultz, Jacob Mincer, and Gary Becker began developing the modern theory of human capital in the 1950s. They compared accumulating new skills and knowledge to acquiring productive physical assets. Acquiring such assets is expensive in time and/or money, and is normally undertaken in anticipation of a future return, whether in money or (for education) personal growth. But these economists and others also identified a crucial difference in the markets for physical and for human capital. Many times, productive investments, such as opening a restaurant, begin with the owner taking on a business loan with a plan to repay the loan out of the investment proceeds. Such loans are usually guaranteed by granting the lender a claim on the restaurant's physical assets, so that the lender can recoup some or all of her investment by selling off the physical assets if the borrower cannot pay. But, unlike a delivery van or a deep-fry cooker, an education cannot be repossessed by a lender. This makes lending money to students for educational investments unusually risky. This is not so much of a problem for people from affluent families, who have other assets to pledge as collateral, but this lack of collateral historically made it essentially impossible for students from low-income families to borrow for college. As a result, economists argued, there is a failure in the private market for human capital: economically sound loans will not get made because lenders cannot protect themselves from default risk. Thus the idea of the federal government guaranteeing loans flows naturally as a way of "perfecting the market." This logic was important in developing the federal loan programs.

Also in the 1960s liberal economists, particularly those associated with the Brookings Institution, became intrigued with the idea of using market tools to achieve liberal ends. Economists such as Charles Schulze (who wrote a book titled *The Public Use of Private Interest* in 1977), Alice Rivlin, and Arthur Okun were skeptical of government bureaucracy and increasingly intrigued with using market-oriented incentives to shape people's choices, rather than having government impose structures. In areas ranging from public utility pricing to armed forces recruiting, policy analysts proposed leveraging market incentives in preference to relying on administrative regulation. This thinking had a good deal to do with federal support flowing directly to students, rather than mainly to institutions. We have a better understanding now of some of the weaknesses of voucher finance, particularly when buyers are ill-informed, but the value of leveraging market incentives in public policy continues to be very widely appreciated.

In reviewing a half-century of federal involvement in higher education finance we can draw two conclusions. First, as clotted and frustrating as the making of federal policy can be, big changes do happen. And second, for better and worse, ideas, including those of social scientists, matter.

50 Years of Progress: Higher Education, Research, and Reform

*Cecilia Elena Rouse*¹

Higher education is at a crossroads. During the past 50 years, more students have gained access to higher education than any period in U.S. history, with the diversity of the student population increasing at an incredible pace. Whether we can continue to improve—or even sustain the success of the past half-century—remains to be seen. I am optimistic, but progress cannot continue without creative thinking, innovation, and vigilance.

Without question, 1965 was a banner year for higher education. President Lyndon B. Johnson launched his “War on Poverty,” a key part of which was the Higher Education Act (HEA). The Act was designed “to strengthen the educational resources of our colleges and universities and to provide financial assistance for students in postsecondary and higher education.”² The year 1965 was also the year that the National Academy of Education (NAEd) was founded, with the express mission of advancing “high quality education research and its use in policy formation and practice.” While it is not clear that the HEA and the NAEd were linked from the outset, during the past 50 years they have been used in tandem

¹ Cecilia Elena Rouse is the Dean of the Woodrow Wilson School of Public and International Affairs and The Lawrence and Shirley Katzman and Lewis and Anna Ernst Professor in the Economics of Education, Princeton University. She was elected to the National Academy of Education in 2010.

² Although there can be nuanced differences, in this essay, I use the terms “postsecondary education” and “higher education” interchangeably.

to improve access to higher education and strengthen postsecondary institutions.

The HEA's original mission was mostly focused on improving *access* to higher education, and most would agree it has largely succeeded in achieving this goal. Whereas in 1960 only about 45 percent of recent high school graduates had enrolled in a 2- or 4-year institution, by 2013 that had increased to about 66 percent (National Center for Education Statistics, 2014). Much of this success can be attributed to Title IV, a key provision of the HEA that created the federal system of financial aid. Title IV established Education Opportunity Grants and the Federal Family Education Loan (FFEL) program in which the federal government began guaranteeing student loans provided by banks and other lenders. The Pell Grant program—the successor to the Basic Educational Opportunity Grant program that was part of the 1972 HEA reauthorization—originally funded only about 176,000 students at a cost of about \$230 million (in 2015 dollars); today there are more than 8 million Pell Grant recipients at a cost of more than \$30 billion (FinAid, 2015; U.S. Department of Education, 2015).

Students use these funds to further their education in a variety of institutions, from programs in community colleges and for-profit institutions to bachelor's degree programs in traditional 4-year colleges. This heterogeneity has opened the door for more diverse students as well. While once a postsecondary student was synonymous with an 18-year-old, unmarried, nonworking individual dependent on parents for financial support, today's postsecondary student is older and often a first-generation college student, with both work and family responsibilities. Nontraditional institutions have developed to accommodate these students with flexible class schedules, diverse academic and occupational programs, and many academic supports. A common question is how much is too much—that is, at what point will too many students enroll in postsecondary institutions? At this point it does not appear we have reached that upper limit, because the economic value of a postsecondary education continues to be robust.

Although the gains in access during the past 50 years have been impressive, they cannot be taken for granted. As is often highlighted in the media, increases in federal financial aid have been outpaced by even larger increases in postsecondary tuition. An oft-quoted citation is that during the past decade college tuition has increased faster than the rate of inflation, perhaps even faster than the cost of health care. Press reports typically focus on the “sticker price”—tuition and fees before accounting for financial aid—of postsecondary education, and note that tuition at public 4-year institutions more than doubled between 1995 and 2015 from \$4,340 to \$9,140 on average (in 2015 dollars).

We need to be careful about these claims, however. While it is certainly true that tuition has increased, and in some cases by large amounts, it is not nearly as bad as reported. Increases in *net* tuition—tuition and fees minus grants and other financial aid—have been much lower. According to The College Board, between 1995 and 2015 net fees and tuition at public 4-year institutions increased by 50 percent, from \$2,020 to \$3,030 (in 2015 dollars), on average. While not as bad as the media has reported, nonetheless this increase came at a time when median household income saw virtually no increase during roughly the same period (U.S. Census Bureau, n.d.). To meet this rise in tuition, therefore, an increasing number of students have borrowed through the federal student loan program. In the mid-1990s, only about one-half of postsecondary students took out student loans, whereas about 70 percent do today. Not surprisingly, the amounts students borrow have also increased; today the average student-borrower graduates from a 4-year college with about \$28,000 in student loan debt. The concern is that the prospect of having to repay so much educational debt is and will continue to discourage able young people from seeking postsecondary education, just at a time when higher-order, analytical skills are so key to a successful career and a productive labor force.

Although accessibility was the main focus of educational reform between 1965 and the mid-1990s, more recently researchers have begun to focus on completion. Only about 60 percent of students who enroll in a 4-year institution complete a bachelor's degree within 6 years. The numbers at 2-year institutions are even starker, with only about one-third of those who enroll completing a certificate, associate, or bachelor's degree. Some of this non-completion reflects the preferences of the students themselves; institutions such as community colleges give them a chance to try out a postsecondary program at a relatively low cost to see whether it is a good fit. But this high rate of noncompletion also stems from ineffective, poor-quality educational programs: uninspiring and/or outdated curricula, inadequate student supports (especially for first-generation college students or those with dependents), lack of institutional coherence (such as complicated and confusing transfer requirements between 2-year and 4-year institutions) or lack of adequate K–12 preparation needed to meet the rigors of postsecondary education.

As we sit at the crossroads of higher education, we need to ask this key question: Is it possible to preserve access, improve student outcomes, and keep college affordable all at the same time? Although I cannot answer definitively, I do know that we cannot figure it out without credible data and research to inform critical decision-making along the way. This means that researchers and organizations such as the NAEd have a key role in helping to identify problems and opportunities, propose sen-

sible solutions, and evaluate what is working and what is not. Fortunately, the past 50 years have also seen an incredible growth in the education research infrastructure. In 1965 only about 1 percent of academic articles focused on education, academic achievement or schooling outcomes; by 2014 that had more than doubled to 2.6 percent.³ In the field of economics alone the number of journals with the word “education” in the title has increased from zero in 1965 to 10 in 2014. One can also see developments on the policy side. The Institute of Education Sciences (IES) was created in 2002 to “provide rigorous and relevant evidence on which to ground education practice and policy and share this information broadly.” One of its signature activities has been the What Works Clearinghouse, a one-stop resource for decision-makers who seek an unbiased and rigorous assessment of the effectiveness of popular education practices, programs, and policies. The Obama Administration has continued this push with such innovative programs as Investing in Education (i3) to promote innovative practices that have a demonstrated impact on student outcomes.

Although not every decision in education needs to be based on the gold standard of a randomized evaluation, the goal is for data, research, and credible evaluation to play a role in the everyday thinking and decision-making of policymakers. Of course this is easier said than done. In the introduction to “Show Me the Evidence: Obama’s Fight for Rigor and Results in Social Policy,” Ron Haskins “guesstimates” that research only accounts for about 1 percent of factors that influence legislation (Haskins & Margolis, 2014). Although its role may seem small, it is not irrelevant. Programs such as Drug Abuse Resistance Education (DARE), Even Start, and Reading First were either cut back or entirely eliminated because of lack of evidence of effectiveness. Another flagship of the War on Poverty, Head Start has been significantly reformed based on credible evidence that the program was not working as effectively as it could be.

Fortunately, all of this activity has spilled over to postsecondary education. A decade ago, MDRC conducted one of the first randomized evaluations of impacts of student interventions at community colleges with its Opening Doors study. Through this study and subsequent expansions, we have learned that “learning communities”—a popular way to organize the first-year of college for students—could be effective when implemented well but that the average program did not have big impacts. Similarly, we have learned that structuring financial aid so that students receive additional payments after meeting certain academic benchmarks can make a small difference to student achievement but alone will not

³ Based on a search of “ProQuest Social Sciences” for “education,” “academic achievement,” and “schooling outcomes” in the title of scholarly journal articles compared to all articles within a calendar year.

dramatically change outcomes. In addition, perhaps most importantly, developmental (or remedial) education is constantly being reviewed and improved based on research; this not only helps the students who need developmental education, but also saves money for state and local governments, which spend billions of dollars each year on these programs.

So I celebrate both the HEA and the NAEd for making it to the half-century. The HEA has played a critical role in building and shaping the U.S. system of higher education, and its importance will only increase as state support for higher education continues to wane and federal investment becomes more critical. However, federal support should not be limited to increased funding for higher education institutions and financial aid for students. Rather, it must include smart policies that seek to improve access and effectiveness. At this juncture, we cannot be complacent, but need to continue identifying weaknesses and opportunities in the system and vigilantly monitor investments to ensure that they are achieving their stated goals. By continuing to fund and support researchers and others attempting to infuse more evidence into policymaking, we will have even more to celebrate when both the HEA and the NAEd turn 100.

References

- FinAid. (2015). Pell grant historical figures. Retrieved from <http://www.finaid.org/educators/pellgrant.phtml>.
- Haskins, R. & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy*. Washington, DC: Brookings Institution Press.
- National Center for Education Statistics. (2014). *2014 digest of education statistics* (Table 302.10). Washington, DC: Author. Retrieved from <https://nces.ed.gov/programs/digest>.
- U.S. Census Bureau. (n.d.). *Historical income data* (Tables H-8). Retrieved from <http://www.census.gov/hhes/www/income/data/historical/index.html>.
- U.S. Department of Education. (2015). *Student financial assistance: Fiscal year 2015 budget request*. Washington, DC: Author. Retrieved from <http://www2.ed.gov/about/overview/budget/budget16/justifications/p-sfa.pdf>.

Affirmative Action and Its Discontents: America's Obsession with Race

Marta Tienda¹

In June 2015 the U.S. Supreme Court announced that it would, yet again, hear the case of *Fisher v. University of Texas*, which for some legal scholars and public intellectuals signals a death knell for affirmative action. The original complaint in Abigail Fisher's name alleged that the university's consideration of race in admission decisions violated the Equal Protection Clause of the 14th Amendment to the U.S. Constitution. Both the U.S. District Court and the U.S. Court of Appeals for the Fifth Circuit ruled in favor of the university. Invoking the 2003 *Grutter v. Bollinger* decision, the Fifth Circuit Court opined that the university's admission system was not unconstitutional and served the educational mission of seeking to garner the pedagogical benefits of diversity. The Supreme Court disagreed and in a 2013 decision remanded the case to the Fifth Circuit court for its failure to apply the "strict scrutiny" standard in rendering its verdict. This may appear to be a legal technicality, but in cases of discrimination and affirmative action jurisprudence, it is profound.

In ordinary circumstances, a Supreme Court ruling would be considered definitive—at least until new evidence or new instances of a violation were brought to bear on the Constitutional issue at hand. That is not the case in the *Fisher* complaint, which the plaintiff's argue hinges on the

¹ Marta Tienda is the Maurice P. During '22 Professor of Demographic Studies and Professor of Sociology and Public Affairs at Princeton University. She was elected to the National Academy of Education in 2015.

ill-defined critical mass concept. The Court's decision to reconsider the *Fisher* case worries supporters of race-sensitive admissions because it will likely revolve around whether the university's narrowly tailored consideration of race in admissions satisfies the strict scrutiny standard; because several justices demonstrated their skepticism of the critical mass concept in oral arguments during the first hearing of the *Fisher* case; and because opponents of affirmative action have filed lawsuits against University of North Carolina and Harvard University that challenge their admission policies. That all three complaints are motivated and financially orchestrated by the conservative Project on Fair Representation (PoFR) rather than specific plaintiffs is irrelevant for the merits of the case.

The Texas and Harvard cases are important for different reasons. The *Fisher* complaint argues that race preferences are unnecessary because the Top 10 percent law, which has been in force since 1998, guarantees admission to all students who graduate at the top of their high school class. In the context of rising levels of school segregation, the law is able to capitalize on segregation to ensure that the two flagship institutions, the University of Texas at Austin (UT) and Texas A&M University (TAMU) attract high-performing minority students. To be sure, the shares of black and Hispanic students at both institutions gradually rebounded at both flagships, but that misses the bigger point about the need for affirmative action for two reasons that also make Texas a propitious target for critics of race-sensitive admissions.

First, a focus on changes in the composition of college campuses is a problematic metric because it presumes that the college-age population neither grew nor became more diverse. Both premises are false. In fact, *enrollment rates* of top-ranked black students admitted to both UT and TAMU were lower under the Top 10 percent regime precisely because growth in the college-age population surged even as successive cohorts of high school graduates became more diverse. Improvements in Texas high school graduation rates have not been matched by higher college enrollment rates, partly because the outsized cohorts of college-eligible students have outstripped the carrying capacity of the higher education system, creating a college squeeze that intensifies competition for slots (Tienda, 2015).

Second, the admission guarantee does not guarantee enrollment. Even if larger numbers of students qualify for admission, enrollment also depends on ability to pay. Application rates of top-ranked students attending schools with large numbers of poor students are notably lower than those of students who graduate from affluent high schools. Both rising tuition and weak financial aid programs dampened college attendance of minority students (Creusere, Fernandez, Fletcher, Klepfer, & Rice, 2014). According to the Center for Public Policy Priorities (2012),

funding for full-time equivalent students has stagnated at 1990 levels in real terms while tuition has quadrupled. These trends effectively restrict college access for low-income students by shifting costs to families with limited ability to pay. A 10 percent cut in the Texas Grants—the major need-based tuition program—in the 2012 and 2013 fiscal years further aggravates the college squeeze for students from low-income families. Both of these developments constrain educational opportunity, but neither triggered moral outrage much less lawsuits.

The Harvard lawsuit is concerning because of its leadership in spearheading individualized, holistic review by conceptualizing diversity as a multi-dimensional phenomenon based on a broad array of applicant attributes, including geography, athletic talent, and extracurricular activities. Race is only one of several factors considered in Harvard's admission plan, and demonstrated academic ability based on high school achievement in addition to standardized test scores. A well-funded advocacy group intent on eliminating race-sensitive admissions throughout the nation, Students for Fair Admissions, filed both the Harvard and University of North Carolina lawsuits on behalf of anonymous applicants. The organization was created to represent prospective Asian Americans applicants, some of whom will be denied admission and thus eligible to claim reverse discrimination on the basis of quotas (Biskupic, 2015).

Two assertions undergird the call to outlaw any consideration of race in college admission decisions. One is that diversity can be achieved using race neutral means. Not only is this claim false for reasons elaborated above, but also percentage plans are not viable for institutions and academic programs that draw on national populations. Even though the Top 10 percent law was a response to a lawsuit filed against the UT law school, the remedy was designed for undergraduate admissions and has never worked for programs, even those at public institutions, that draw on national applicant pools. This argument is facile at its core but nonetheless is not applicable to graduate and profession programs throughout the nation. The second assertion behind the complaints is that race preferences have the unintended consequence of discriminating against Asians, whose admission rates lag those of other groups with lower average credentials. There is no mention in the lawsuit against Harvard that Asian students were the major beneficiaries of the Texas Top 10 percent law. The case is not about Asian students, who represent the latest pawn in a political quest to end affirmative action in college admissions throughout the land.

In addition to wasting enormous resources for litigation, the continued legal drama over affirmative action has two unfortunate unintended consequences. First, it deflects attention from the broader social goal of campus diversification, namely integration in order to harness the peda-

gological benefits of diversity through vigorous intellectual exchange and collective problem solving. Sandra Day O'Connor's admonition in *Grutter* that the nation's public institutions "should be pursuing the larger national project of integration" is totally lost in the frenzy of legal defense and offense in the latest round of litigation before the high court. The back-up plans at Harvard and the University of North Carolina virtually ensure that the mission to end race-sensitive admissions will persist and ultimately prevail.

The consequences of doing so are well documented—at least in terms of enrollment at selective institutions. Recently *The New York Times* (2015) published an infographic that illustrates the consequences of legal and statutory bans on affirmative action. In California and Texas, for example, the state's public flagships enrolled fewer black and Hispanic students compared to the period prior to the bans. These worrisome trends transcend philosophy, ideology, and politics because they are occurring at a time of major population diversification and global international competition. Investing in the education of the nation's burgeoning minority populations is a compelling state interest, as reasoned in the *Grutter* decision in order to ensure that the pathways to leadership are visibly open. Diversification is not a transitory feature of the United States; however, the fate of affirmative action in the courts will determine whether and how much group membership becomes an enduring social division in the years to come. Rather than focusing on what constitutes a critical mass, legal teams should instead underscore the state's compelling interest in protecting narrowly tailored racial preferences in college admissions. The nation's future prosperity depends on broadening the pathways to leadership and garnering the myriad economic and social benefits from diverse college campuses.

References

- Biskupic, J. (2015). Special report: Behind Harvard case over Asian admissions, a broader agenda. *Reuters*. Retrieved from <http://www.reuters.com/article/2015/06/08/usa-harvard-discrimination-idUSL1N0YU14M20150608>.
- Center for Public Policy Priorities. (2012). *Texas investment in higher education lags behind student needs and workforce demands*. Austin, TX: Author.
- Creusere, M., Fernandez, C., Fletcher, C., Klepfer, K., & Rice, E. (2014). *State of student aid and higher education in Texas*. Round Rock, TX: TG Research and Analytical Services.
- Fisher v. University of Texas at Austin*, 133 S.Ct. 2411 (2013).
- Grutter v. Bollinger*, 539 U.S. 306 (2003).
- The New York Times*. (2015). How minorities have fared in states with affirmative action bans. Retrieved from http://www.nytimes.com/interactive/2013/06/24/us/affirmative-action-bans.html?_r=0.
- Tienda, M. (2015). Texas' educational challenge: A demographic dividend or bust? In P. M. Orrenius, J. Cañas, and M. Weiss (Eds.), *Ten-gallon economy sizing up economic growth in Texas* (Chapter 5). New York: Palgrave Macmillan.

The Higher Education Act of 1965: A Half-Century's Worth of Contrasting Public Philosophies and Controversies

*William G. Tierney*¹

In the depths of the Civil War, Congress passed, and the President signed into law, the Morrill Land Grant Act. At the time the law was the most significant piece of federal legislation pertaining to higher education in the young country's history. Even as President Lincoln struggled to keep the union together he viewed education as the great equalizer. The law, passed in 1862, set the country on a course for the creation and support of public higher education (Thelin, 2011). A second land grant act, passed in 1890 that included African Americans, only further solidified the role of the federal government in higher education policy. Although different states developed different systems of public higher education, it is fair to suggest that without the vigorous support of federal legislation, the United States would not have developed what has come to be seen as the strongest system of higher education in the world.

The Serviceman's Readjustment Act of 1944, better known as the GI Bill, was the 20th century's first major act that revolutionized American higher education by expanding college-going to returning veterans. Other legislation followed such as the National Defense Education Act, which created aid for students studying science, technology, engineering, and mathematics (STEM) and foreign languages as a response to the Soviet Union's launching of Sputnik (Graham, 1984). In 1947 the Truman Com-

¹ William G. Tierney is the Wilbur-Kieffer Professor of Higher Education at the Pullias Center for Higher Education, University of Southern California. He was elected to the National Academy of Education in 2014.

mission called for an expansive federal role in higher education, which further cemented the idea that Congress had a role to play in designing an equitable postsecondary system throughout the country.

The Higher Education Act (HEA) of 1965 is a direct descendent of the belief that the federal government had not only a role, but also the ability to stimulate educational opportunity. This ambitious piece of legislation sought to increase the economic and social well-being of the country by expanding participation in higher education (Hearn, 1993). To be sure, the act was in keeping with other Great Society legislation of the time such as the Civil Rights Act and the Voting Rights Act. What differed with the HEA, however, was the philosophic assertion that the government should be at the forefront in the attempt to expand educational opportunity via public support for higher education.

When one looks back a half-century, much less a century and a half when the Morrill Act was passed, the confidence and ability of Congress to legislate such dramatic efforts aimed at educational opportunity offers a stark contrast to the gridlock that currently exists in Washington, DC. Indeed, the inability for the current Congress to pass legislation not only highlights partisan divides but also a philosophic resistance to the tenets of legislation such as the HEA. As opposed to an assumption that government could be a force for economic and social good through education, the overriding public philosophy today is that government should step back and let the market decide. Individual determination, rather than governmental intrusion, is seen as the best path to increasing economic growth. Public higher education, once seen as a quasi-public good, today is viewed more as a private good that individuals should pay for because they will be the primary beneficiaries.

The HEA is less known as a cohesive piece of legislation and better understood for the various "Titles" that are parts of the legislation, and for the reauthorizations that have occurred since its inception. The changes to the HEA are an example par excellence of "an incremental and fragmented federal policy-making process" (Hannah, 1996, p. 498). Strategies and beliefs have swung wildly in support of one public policy and away from another only to swing back again. In what follows I consider five distinct parts of the legislation that underscore the assumptions about the government's role in advancing educational opportunity and how perceptions have changed over time.

Developing Institutions

The second Land Grant Act tried to establish "separate but equal" colleges for historically disenfranchised groups, which at the time meant African Americans. Historically Black Colleges and Universities (HBCUs)

were the outcome of land grants. Title III initially provided support for HBCUs, community colleges, and liberal arts colleges, which all were seen as out of the mainstream. Over time, however, the bulk of Title III funding has supported HBCUs and also tribally controlled colleges for Native Americans and Hispanic-serving institutions for Latinos. Title III provides aid to these sorts of institutions, but the aid has waxed and waned since the law's inception.

At the law's outset the assumption was that racial and ethnic minority institutions were part of a postsecondary patchwork quilt that helped increase economic opportunity for low-income people of color. The legislation provided a few million dollars for each HBCU that was supposed to be buttressed with state support and private philanthropy. Throughout their history, however, HBCUs and the other related institutions have been under-funded, especially so during the last recession (Gasman, 2010). A long-standing criticism of this section of the HEA is that it is supporting racially defined institutions that frequently under-perform. Fifty years ago a significant majority of African Americans attended and graduated from HBCUs, whereas today it is only 10 percent; therefore, some allege that these institutions have outlived their usefulness. A counter-response has been that HBCUs graduate the lion's share of students of color in many fields and that if they were funded at the level of their historically white counterparts, then performance would increase. Critics, however, point out that over the past decade HBCUs have received close to \$3 billion in support, and the outcomes of that support are unclear.

Student Assistance: Grants Versus Loans

Perhaps the most important part and the most discussed aspect of the HEA is that which allows low-income and middle-class students to make use of grants and loans to pay for their education. Title IV's programs have been called "the most far-reaching as well as the most controversial" aspect of the bill (Cervantes et al., 2005, p. 20). Although previous legislation provided funding for specific students (e.g., returning veterans) or particular areas of study such as science and math, Title IV opened the gates for federal funding of student attendance at college. In the 2013–2014 academic year Title IV provided \$169 billion in aid—a 100 percent increase over a decade (Burke, 2014).

Originally, grants were intended for low-income students and loans for the middle class. Over time, however, although the amount of grants has increased, they have not kept pace with the cost of tuition. Loans now account for a larger payout to students. In the reauthorization of 1972 what has come to be known as "Pell Grants," after Senator Claiborne Pell, became a substantial way to fund a student's costs for college (Curs,

Singell, & Waddell, 2007, p. 284). As I touch on below, the type of institution that is able to receive such funding has become a point of contention. More importantly, whether a student should be entitled to receive grants from the government has become a topic for debate. One argument is that the taxpayer should not be paying for individuals to gain a college education. A related argument is that when aid increases postsecondary institutions increase their tuition. Hence, a decrease in support to students is likely to decrease the cost of college.

College Access Programs

The HEA has long struggled with finding programs that will increase access to postsecondary institutions. What have come to be known as the TRIO programs (Upward Bound, Talent Search, and Student Support Services) have been increased to include other programs such as GEAR-UP during subsequent reauthorizations. Although the programs are well-intended ideas aimed at increasing access, they have proven to be controversial. Various evaluations have pointed out negligible findings, which in turn have created a firestorm of criticism pertaining to the methodologies employed and the political nature of the studies. Various administrations have tried to eliminate or curtail the funding for the programs, but TRIO and GEAR-up endure.

For-Profit Higher Education

Although various issues have been controversial in the HEA, perhaps none has been more so than the ability of for-profit colleges and universities (FPCUs) to have students who make use of federal dollars. Private universities have always been able to have students make use of federal grants and loans but they were nonprofit. In 1972 Congress amended the HEA to allow FPCUs to access federal student aid (Tierney & Hentschke, 2007). Two conditions of accessing aid were that the institution had to be accredited and that at least 15 percent of an institution's revenue needed to come from non-federal aid. Subsequently, this provision was changed to the "90-10 rule" where 10 percent of revenue needed to come from income other than federal and state grants and loans. One of the unintended consequences of federal aid in general, and the inclusion of FPCUs in particular, was to dramatically expand the recipients of aid from traditionally 8- to 22-year-old students to undergraduates older than 22 years (Wolanin, 1998).

The concerns about FPCUs were threefold. First, many have argued that profit-making institutions are anathema to legislation that looked on education as a public good. Second, others argued that the quality of

the institutions led to students saddled with far too much debt and little chance of gainful employment. Third, for-profit institutions were claimed to have targeted naïve first-generation consumers by predatory practices that enabled students to access loans, but not understand what they were really doing. Proponents of FPCUs countered that they were tax-paying institutions that were trying to meet consumer needs; if they were not providing a good product, then the market would not allow them to continue (Coleman & Vedder, 2008).

Accreditation

A current controversy pertains to accreditation, and in turn, institutional rankings. Senator Alexander of Tennessee has questioned whether accreditation is meeting the needs of Congress (Alexander, n.d.). The HEA deemed that an institution needed to be accredited to be eligible for federal aid. The for-profit world has looked on accreditation as a members' "cartel," which keeps out institutions with different practices (such as online learning). Other critics have assailed accreditation as a conundrum insofar as it accredits institutions that subsequently encounter serious problems that force them to go out of business. If accreditation is supposed to improve quality, the critics claim, then there is little evidence that it has done that during the past half-century. A suitable measure to gauge quality has proven elusive, and whether the government should develop and/or publish rankings of institutions has been subject to vigorous debate.

Conclusion

The HEA and its multiple provisions have gone through various permutations as Congress's philosophies have changed and evolved. Although there is virtually no discussion about repealing the legislation, it is fair to say that such an Act as it is currently configured could not get through Congress today. The country has changed, and the assumption that higher education is a public good that should be largely funded by the federal and state governments is no longer a viable option. The assumption is that the individual should shoulder a greater share of the cost for his or her education, and that even the expansion of higher education is of questionable value. Such a viewpoint stands in stark contrast to the time of President Johnson, who saw education as the "great equalizer." Although the HEA has had various measures in need of repair and reform, a judicious observer could certainly conclude that it has had an impact—more individuals have been able to access higher education because of the HEA. The question remains, then, if the country still has a

public philosophy that maintains equal opportunity is a cherished ideal, then what levers might be employed on its behalf?

References

- Alexander, L. (n.d.). Higher education accreditation concepts and proposals. Senate committee on health, education, labor & pensions. Retrieved from <http://www.help.senate.gov/imo/media/Accreditation.pdf>.
- Burke, L. M. (2014). Reauthorizing the higher education act—Toward policies that increase access and lower costs. *The Heritage Foundation Backgrounder*, No. 2941. Retrieved from <http://report.heritage.org/bg2941>.
- Cervantes, A., Creusere, M., McMillion, R., McQueen, C., Short, M., Steiner, M., & Webster, J. (2005). *Opening the doors to higher education: Perspectives on the Higher Education Act 40 years later*. San Antonio, TX: Texas Guaranteed Student Loan Corporation.
- Coleman, J., & Vedder, R. (2008). *For-profit education in the United States: A primer*. Washington, DC: The Center for College Affordability and Productivity. Retrieved from <http://files.eric.ed.gov/fulltext/ED536281.pdf>.
- Curs, B. R., Singell, L. D., & Waddell, G. R. (2007). The Pell program at thirty years. In *Higher education: Handbook of theory and research* (pp. 281–334). Netherlands: Springer.
- Gasman, M. (2010). Comprehensive funding approaches for historically Black colleges and universities. In *HBCU symposium: Setting the agenda for Historically Black Colleges and Universities*. Durham, NC: North Carolina Central University.
- Graham, H. D. (1984). *The uncertain triumph: Federal education policy in the Kennedy and Johnson years*. Chapel Hill, NC: University of North Carolina Press.
- Hannah, S. B. (1996). The Higher Education Act of 1992: Skills, constraints, and the politics of higher education. *The Journal of Higher Education*, 67(5), 498–527.
- Hearn, J. C. (1993). The paradox of growth in federal aid for college students, 1965–1990. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (vol. 9). New York: Agathon Press.
- Thelin, J. R. (2011). *A history of higher education* (2nd ed.). Baltimore, MD: Johns Hopkins University Press.
- Tierney, W. G., & Hentschke, G. C. (2007). *New players, different game: Understanding the rise of for profit colleges and universities*. Baltimore, MD: Johns Hopkins University Press.
- Wolanin, T. (1998). Pell grants: A 25 year history. In L. E. Gladieux, B. Astor, & W. S. Swail (Eds.), *Memory, reason, imagination: A quarter century of Pell Grants*. New York: The College Entrance Examination Board.

The Interplay Between Education Research and Policy

Education's Double Helix

*Eric A. Hanushek*¹

The scientific study of education faces challenges that are more acute than those found in most other scientific endeavors. Educational research feeds directly into policymaking and into changes in the way our schools operate, which influences not only the topics considered in research but also at times the research itself. In other fields of research regular linkage to widespread policies is not nearly so common. This linkage in education research puts extra pressure on researchers, but it also confers extra benefits. The National Academy of Education (NAEd) is in my opinion an institution that should not only ensure the highest level of development of the scientific study of education but also oversee the appropriate use of educational research in the policy sphere.

The connection of research and policy appears quite broadly across education, and it is more difficult to find areas of active education research that are removed from this association than fit into it. The ongoing discussions of the common core curriculum, test-based accountability, use of technology in the classroom, charters and school choice, school desegregation, professional development programs, early childhood programs, and teacher preservice training all illustrate the regular interaction of major streams of research with K–12 policy. In addition, while not emphasized here, a similar set of topics ranging from access to debt burdens to for-profit schools fill the higher education space.

¹ Eric A. Hanushek is the Paul and Jean Hanna Senior Fellow at Stanford University. He was elected to the National Academy of Education in 2006.

Many of the general themes have been present for a long period of time, but they have changed in character over the past half-century. Past policy discussions were supported more by philosophy, opinion, and assertion than by direct evidence. However, a change occurred, and it picked up momentum. Generally accepted scientific principles became more the norm for education research. Scientific evidence was introduced into the policy mix, and increasingly scientific evidence became demanded in policy discussions.

The example that I am drawn to for illustrating the evolution of the research-policy linkage, perhaps because it has occupied a substantial portion of my own research, is the investigation of teacher effectiveness. The importance of the teacher has certainly always been known, but policies related to teachers have evolved to reflect the growing scientific research on effectiveness.

The ubiquitous single salary schedule that rewards teachers on the basis of experience and graduate education had been the result of searching for objective criteria by which to reward teachers while removing the possibility of inappropriate subjective judgments. This structure of salaries made intuitive sense, because most occupations see improvement in individual productivity with experience and it is difficult to argue against the value of more schooling for teachers because providing useful skills through schooling is precisely the job of teachers.

However, as evidence started to become available, there was a growing recognition that experience and teacher education might not be closely related to the effectiveness of the teacher in the classroom. Specifically, these pay parameters were not closely related to student outcomes.

At the same time another set of studies investigated whether there were different learning gains by students across classrooms. These studies, now commonly labeled value-added studies, uniformly found that the differential achievement gains of students depending on specific teacher assignment were enormous and could have lasting impacts on students.

The underlying research, like most scientific endeavors, has evolved. Researchers from a variety of disciplines have contributed to expanding and to refining the analysis. Indeed the evidence has become more consistent and reliable with new investigations.

It is useful, however, to step back so as to follow the interaction of this research with policy discussions. When these observations about teacher effectiveness are combined, one of the most controversial research conclusions enters into the debates. If teachers are paid according to factors unrelated to classroom achievement, if personnel costs make up the vast majority of overall spending in schools, and if there are large differences in learning across teachers, then spending on schools might not be consis-

tently related to student outcomes. Although the subject of considerable debate, the scientific research generally points to such an inconsistency. Here is where the linkage of research and policy blossoms.

On the overall spending results, it is clear that many researchers have their own views on the appropriate amount of spending for schools and specifically on whether spending should expand, particularly to help disadvantaged students. It takes considerable effort to separate the scientific research from the particular policy preferences of researchers—and unfortunately it does not always happen.

What are the correct policies with respect to teachers that follow from this line of research into teacher effectiveness? Obviously there are deep divisions in people's ideas about the relationship of these research findings with teacher policy. Some of the divisions result simply from personal perspectives or institutional pressures to advocate a continuation of current policies, independent of any scientific evidence. Advocacy positions are plentiful enough that specific examples are unnecessary.

However, the scientific evidence on variations in teacher effectiveness does not yield specific policies. They suggest (to me) that there are large gains to be made from focusing attention on improving the effectiveness of all teachers, and particularly the teachers of disadvantaged students. How to translate this into policy, however, is not obvious. Nor does the scientific research provide clear guidance.

To be sure, I personally have some opinions about the best ways to proceed. These opinions are informed by the scientific evidence, but they also go beyond the existing evidence. As such, it is incumbent on the scientist to make clear where the science ends and where the policy opinions begin.

One common extension to this separation problem needs to be considered explicitly. Scientific studies, particularly in the early phases of research into a specific area, often come up with quite different results. Too commonly, people with an advocacy viewpoint will collect a number of studies that have the results that support their position and then report them as scientific proof of their position. It takes little thought to understand that this approach is not scientific, even if the underlying data points of the advocacy position come from rigorous scientific studies. Simply put, no scientific methodology calls for selecting evidence on the basis of the answers in given studies.

This account of education's double helix—the intertwining of scientific research and educational policy—leads me to suggest a more active future role for the NAEd. The NAEd can and should help to clarify the role of scientific research into education. The future evolution and improvement of education depends importantly on getting the science

right. Here the NAEd, composed of the best scientists in the area, should take the lead and help policymakers and the public understand the separate strands of research and of policy. Even though much of the scientific research has policy implications, there is quite generally a distinction between the science and the policy.

Research and Policy: The Need to Tie the Knot

*Jack Jennings*¹

In the United States, the importance of education is axiomatic. Presidents and other office holders, business people and civic leaders, and countless others acknowledge the necessity of having an educated citizenry and workforce.

Because of education's significance, many efforts were undertaken, especially in the past half-century, to improve the public elementary and secondary schools. From my perspective, what was missing in many of these "reforms" was a basis of evidence that they would succeed.

In other words, the country needs good researchers who can find the most effective ways of raising the educational level of the American population. In my involvement in education policy for nearly 50 years, I can see improvement in the quality of both educational researchers and their work; however, more remains to be done to continue to improve education research.

The other part of the equation, though, is that *much more* must be done to connect what is learned through education research to the policy-making process. To state it differently, the politicians should use what is learned through education research to enact sound legislation and to make other policies that affect the schools. A related necessity is that those who hold public office should use this information correctly; they should understand it well enough not to misuse the findings.

¹ Jack Jennings is the Founder and Former President and CEO of the Center on Education Policy. He was elected to the National Academy of Education in 2011.

Considering these general conclusions, the National Academy of Education (NAEd) has several roles to play in this process. First, it is prestigious to belong to the Academy, a selective organization admitting members based on their accomplishments. That function is important to assist in raising the prestige and quality of education research. Acknowledgment by one's peers is an encouragement to continue to do good work. In addition, meeting with other researchers of high quality and accomplishment is a way to learn more and therefore do better work.

The second aspect that I mentioned earlier is the need to connect policymakers to researchers, or to state it differently, to find ways that sound research findings can be used to make good policy. I find this aspect missing in the NAEd. It is a sorely needed addition. I will provide two examples of how research did not lead to good policy.

In California, the state enacted a statute that reduced class size for the early grades. Governor Wilson based that proposal on research that showed that students from low-income families gained educationally if they were in classes of fewer students than normal. The governor, though, made that a generalization for all students. That policy had the reverse effect of what was intended. Because all school districts had to reduce the size of early grades, teachers shifted from districts with large numbers of students from low-income families to those with more affluent families. Those latter districts attracted those teachers because they could offer higher pay and better working conditions. Students in schools with concentrations of poor students suffered the loss of those experienced teachers.

The second example is the No Child Left Behind Act of 2001 (NCLB). That federal law imposed penalties on schools and districts that did not raise the test scores of sufficient numbers of students to reach state targets for moving all students to proficiency in mathematics and reading by 2014. To my knowledge, there was no sound research to back the premise of that law, which was that punitive measures for not raising student test scores would result in an increase in student achievement. The lack of that basis has led to broad opposition to NCLB, and the law is now being repealed section by section through waivers from its provisions to the states by the U.S. Secretary of Education and through legislation working its way through Congress.

Former Congressman John Brademas and I are among the only members of the several hundred that belong to the Academy who have been intimately involved over the years in establishing policies for education. I would hope that there would be others who would be selected for admission.

The Academy's membership consists of many of the best education researchers in the country. It lacks members who know how to incor-

porate into policies what those researchers have learned. I hope in the second half-century of the Academy's existence that aspect of education research is addressed and remedied.

In summary, what is the benefit of knowing something if that knowledge is not used to make education better? The NAEd should have a role in both encouraging good research and having it used to govern our society.

Melding Political Sustainability Analysis with Education Research

Lorraine M. McDonnell¹

The 50th anniversary of the Elementary and Secondary Education Act of 1965 (ESEA) also commemorates a larger milestone marking the enactment of the War on Poverty programs. One by-product of that unique period in the history of U.S. social policy was the beginning of program evaluation as a scientific enterprise. Among its applications have been numerous studies assessing whether or not education policies are producing their intended effects, and ascertaining the factors that account for their identified outcomes. In parallel with the development of program evaluation as an applied social science field was the emergence of implementation research focused on determining the extent to which policies are altered as they move through the federal system and explaining local responses to top-down policy incentives and mandates. Together these analytical approaches have been used over several decades to produce a robust body of research depicting not only the outcomes associated with different education policies, but also the process that led to those outcomes and the factors explaining them. That knowledge can then be used to suggest how policies might be modified to be made more effective.

However, knowing how a policy might be altered to increase its educational effectiveness is a necessary condition for change, but not a sufficient one. The sufficient condition relates to the politics of a policy. Those politics include not only who supports and opposes the status

¹ Lorraine M. McDonnell is a Professor at the Department of Political Science, University of California, Santa Barbara. She was elected to the National Academy of Education in 2009.

quo or proposed alternatives, but more importantly, how those political dynamics have become embedded in the policy. They make some elements highly resistant to change through policy feedback that reinforces them, while other parts may be more amenable to change. In these cases, support may come from negative feedback from those adversely affected by the status quo. Change may also be prompted by the emergence of interests with new definitions of policy problems and their solutions. Political scientists and policy analysts have long considered political factors such as interest group and partisan preferences in explaining why particular proposals get on decision-makers' agendas, and the reasons why some are enacted and others are not. More recently, however, political scientists who study public policy have paid greater attention to the political dynamics that develop after a policy's enactment.

This essay outlines how analyses of the political sustainability of education policies can complement research on their effectiveness in promoting student learning. It describes two theoretical concepts now common in the analysis of other domestic policy domains. It then applies them in a brief history of the testing provisions in ESEA Title I, focusing on elements that have remained stable over its 50-year evolution and ones that have changed. Although space limitations only allow for considering the ESEA testing provisions as an illustrative example, the purpose of a more in depth exercise would be to identify those elements of a policy that are most amenable to change and those likely to be politically difficult to alter.

Assessing Political Sustainability and Potential for Policy Change

Two theoretical concepts have been particularly important in developing an understanding of the political sustainability of policies. The first is policy feedback, an analytical framework for examining how a policy's initial design shapes political responses to it and, in turn, influences its sustainability and that of subsequent policies that may develop (Hacker, 2002; Mettler & Soss, 2004; Skocpol, 1992). The concept is based on the assumption that policies establish new institutional rules and structures or augment existing ones. These rules specify the conditions under which policies allocate benefits to some individuals and groups and impose costs on others. For interest groups and other political actors, these costs and benefits create incentives for mobilizing to protect their benefits or minimize their costs. At the level of the mass public, policies can generate interpretative effects as individuals come to understand how costs and benefits affect them personally, and infer what that signals about their status as citizens (Pierson, 1993). These mobilization and interpretative effects create political dynamics that constitute feedback and shape future policy. Depending on whether the feedback is positive or negative,

it can reinforce existing policy, or it may lead to changes in a policy's key elements.

The second concept is the notion of path dependencies, which refer to elements of a policy that become difficult to reverse. Often they result from seemingly small decisions that set a policy on a trajectory that becomes increasingly difficult to alter because the political and other costs of switching to a different path or policy option increase over time (Pierson, 2000). Path dependencies are created by policymakers' decisions, some of which are prompted by positive policy feedback generated through interest group mobilization. However, path dependencies are also reinforced by enduring characteristics of the U.S. political system. Examples include federalism with authority and influence shared among multiple governmental levels, and the separation of powers that produces different types of policies depending on whether they originate from courts, legislatures, or executive agencies.

Whether path dependencies result solely from the effects of policy feedback or in combination with institutional characteristics, they can become integral to a policy, shaping its overall sustainability and that of its major elements (e.g., funding formulas, eligibility requirements). Because policies are rarely completely reversed, attention to policy feedback and path dependencies can inform analysts in identifying which parts of policies are likely to persist and which might be altered or modified. Ideally, the results of applying these analytical tools systematically to identify points of leverage for policy change should be predictive to be useful. However, the starting point for such analyses has to be retrospective. One common approach used by political scientists interested in the political outcomes of policies once they are enacted is theoretically grounded historical analysis. Such analyses have examined a variety of domestic policies ranging from agricultural subsidies to health care (e.g., Morgan & Campbell, 2011; Patashnik, 2008; Grogan & Patashnik, 2003). However, few of these political historical analyses have incorporated the results of either implementation or evaluation studies documenting how well policies have performed in meeting their instrumental goals.

Researchers studying the politics of education policy have tended to focus on the factors shaping the initial passage of legislation. They have only recently begun to recognize the importance of systematic analyses of the politics that develop over a policy's historical development for informing consideration of future policy. However, even though they have come late to political sustainability analysis, education policy researchers have an advantage over their peers in other policy areas. They can draw from a rich body of research on the educational effectiveness of the policies they study, and many work in interdisciplinary environments that can facilitate the melding of educational and policy research.

ESEA Title I Testing: A Brief Illustration of Assessing Political Sustainability²

The longevity of ESEA and the testing provisions embodied in it suggests that this policy's political sustainability is well-established. Nevertheless, a politically sustainable policy is not necessarily a static one, and in the case of ESEA, some elements have changed in significant ways over its history. Consequently, the analytical task is twofold: identifying which policy elements are stable and which have changed, and then explaining the reasons for those differences. Based on that retrospective analysis, it should be possible to predict with some certainty where sources of stability and change will lie in the future.

Title I requirements specifying which students should be tested, how they should be assessed, and how the results should be used have served two purposes as a policy tool and a measurement instrument. Over its history, these purposes have been defined by a stable core of requirements that have included (1) a focus on student assessment as central to an evaluation and accountability strategy; (2) testing as a tool to leverage state and local practice; and (3) a constituency with testing and evaluation as part of its advocacy strategy even though its membership has changed over time. The role of testing as a central component of Title I is an example of a path dependency that began with a seemingly small event: Senator Robert Kennedy's ultimatum that his support for the original ESEA legislation was contingent on the inclusion of a requirement to hold educators responsible for student achievement as a major criterion in judging Title I's effectiveness. Similarly, because other language in the initial statute—still applicable today—prevents the federal government from exercising direction over local curriculum and instructional programs, it has to rely on requiring *ex post* reporting of program results as the primary tool for leveraging state and local practice.

The result has been positive policy feedback stemming from the creation of a vast system of state and local testing, and reinforced by a coalition of interest groups that view the testing requirements as an accountability mechanism to monitor whether local Title I programs are serving their intended beneficiaries effectively. Although the membership of this coalition has shifted throughout ESEA's history, what has been constant is that these groups have supported enforcement of Title I's categorical requirements in counterpoise to traditional education interest groups that have advocated for greater state and local flexibility in program administration.

Although its original inclusion in Title I can be explained by Kennedy's

² This section is excerpted from McDonnell (in press).

amendment, the stability of testing as a key policy tool for federal leverage is best explained by the institutional factors that define education policy in the U.S. federal system—namely, the federal government’s limited formal authority and an ingrained political culture legitimating state and local autonomy along with the variation it produces. The centrality of testing requirements in Title I is a case of strong path dependency where institutional characteristics fundamental to the nature of the American state have made the costs of diversion from that path politically and administratively prohibitive. However, as policy ideas, testing technology, and political dynamics have shifted, the configuration and direction of that path have also been altered.

As a result, within a stable core of testing requirements, three elements of the policy have significantly changed: (1) the focus of accountability has moved from monitoring the distribution of inputs to evaluating the effectiveness of program outcomes; (2) states are now required to incorporate Title I recipients into their standards and assessment systems as they apply to all students; and (3) the technical characteristics of tests have changed and their uses have become more consequential. The same institutional and interest-based factors that explain the stable elements of ESEA’s testing provisions also explain these changes.

Federalism and the preeminence of states in K–12 education constrain the federal government, thus leading to a greater reliance on testing and fiscal reporting requirements as proxy indicators—albeit imperfect ones—to enforce ESEA’s categorical program goals of moving states and localities toward improved learning opportunities for low-income students. The 50-year history of ESEA suggests that a wholly different strategy is not likely to be politically or administratively feasible. At the same time that federalism has functioned as an institutional constraint ensuring the stability of the testing requirements, it has also provided the federal government with opportunities for changes that have extended and strengthened its programmatic reach over state and local behavior. In the wake of the *Nation at Risk* report and the implementation of standards-based reforms (SBRs) in a number of states, academic content and performance standards along with standardized assessments became an integral part of state policy. This major development at the level of the governmental system with constitutional responsibility for education allowed the federal government “to borrow strength”—taking advantage of the arguments that states had already made to justify involvement of higher levels of government in classroom processes and outcomes, and the capacity of administrative structures that state reforms had already created (Manna, 2006).

Just as the federal government borrowed strength from the states, groups with an interest in testing as part of a reform strategy have taken

advantage of the multiple policy arenas in the federal system in advancing their agenda. Their promotion of SBRs at all governmental levels has resulted not only in a range of policies, but also in the establishment of new institutions to develop and maintain those reforms. The testing infrastructures now operating in states and local districts and used in implementing the requirements of No Child Left Behind (NCLB) are prime examples. Newer entrants into the education policy arena, such as business interests and a range of groups promoting various reform agendas, actively support the NCLB testing requirements as part of a larger accountability strategy. Their support has ensured the continuation of testing as a policy tool central to Title I, but it has also helped promote the changes that have made the requirements more prescriptive.

This brief overview of the evolution of ESEA's testing requirements has several implications for assessing the possibility for future change. First, for a well-established policy such as ESEA, analyzing political sustainability and the prospects for modifying it requires disaggregating the policy into its key components such as the testing provisions. Second, the initial decision that made student testing a part of ESEA was largely due to political expedience to ensure its passage. However, the federal government's need to mitigate the constraints of federalism, reinforced by policy feedback generated by a coalition of interests invested in testing, ensured that it would remain part of ESEA's stable core. The inclusion of mandated testing in current reauthorization bills suggests that this path dependency is likely to endure. At the same time, a third implication highlights the potential for change within a stable core. That possibility emerges when engaged interests create a form of negative policy feedback by proposing new definitions of the policy problem and different approaches to solutions. In the case of ESEA, SBRs were viewed as a solution to Title I's documented shortcomings as well as to the perceived failure more generally of the U.S. education system in offering high-quality and equitable learning opportunities to all students. Just as SBRs strengthened the requirement to assess students by focusing on how those results are used, the levers for future change in ESEA are likely to lie in test use rather than in the front-end provisions to test students on a regular basis.

Conclusion

Over the past several decades, there has been a growing recognition that politics cannot be divorced from education. Yet politics is often viewed as a largely external factor to be considered only at the time of policy enactment or if strong opposition emerges. In contrast, politics should be seen as an integral characteristic of policy that develops over its history, shaping its operations and the likelihood of change. Because

of its interdisciplinary membership and its mission to enhance the study of education from multiple perspectives, the National Academy of Education is in a unique position to encourage the melding of evaluation research on educational effectiveness with historical and political sustainability analyses.

References

- Grogan, C., & Pastashnik, E. (2003). Between welfare medicine and mainstream entitlement: Medicaid at the political crossroads. *Journal of Health Politics, Policy and Law*, 28(5), 821–858.
- Hacker, J. S. (2002). *The divided welfare state*. New York: Cambridge University Press.
- Manna, P. (2006). *School's in*. Washington, DC: Georgetown University Press.
- McDonnell, L. M. (in press). Stability and change in Title I testing policy. *Russell Sage Foundation Journal of the Social Sciences*.
- Mettler, S., & Soss, J. (2004). The consequences of public policy for democratic citizenship: Bridging policy studies and mass politics. *Perspectives on Politics*, 2(1), 55–73.
- Morgan, K. J., & Campbell, A. L. (2011). *The delegated welfare state*. New York: Oxford University Press.
- Patashnik, E. M. (2008). *Reforms at risk: What happens after major policy changes are enacted*. Princeton, NJ: Princeton University Press.
- Pierson, P. (1993). When effect becomes cause: Policy feedback and political change. *World Politics*, 45, 595–628.
- Pierson, P. (2000). Increasing returns, path dependence, and the study of politics. *American Political Science Review*, 94(2), 251–267.
- Skocpol, T. (1992). *Protecting soldiers and mothers: The political origins of social policy in the United States*. Cambridge, MA: Belknap.

The National Academy of Education's Role in Bridging Research and Practice

Lauren Resnick¹

I was elected to National Academy of Education (NAEd) membership in 1981. Over the intervening decades, I have attended most of the NAEd's meetings, participating in its debates and sometimes organizing them, offering a presentation or a commentary, and enjoying the lively discussions.

As I look back now, I can see that I was rather more active in the *practice* and *policymaking* side of education than most other members. This seems odd in retrospect, because I have been a university faculty member continuously from 1966 to the present. From my comfortable "perch" at the University of Pittsburgh, and most significantly its Learning Research and Development Center (founded in 1963 by Robert Glaser, also a distinguished founding member of the NAEd), I moved back and forth between scholarship and policy/practice without discomfort or skepticism on the part of my NAEd colleagues.

I benefited both from the formal presentations and the side discussions. Conversations became even more lively and interesting when the NAEd entered into partnership with the Spencer Foundation to create the NAEd/Spencer Dissertation and Postdoctoral Fellowship Programs and we had younger scholars constantly with us. The intense and multi-vocal discussions of important issues of education policy and practice kept me going to meetings even when my calendar was crammed.

¹ Lauren Resnick is Distinguished University Professor at the University of Pittsburgh. She was elected to the National Academy of Education in 1981.

I did not even know that the NAEd existed when I was first invited by leading NAEd scholars John B. Carroll and Jeanne Chall to prepare a paper for a volume they were editing (*Toward a Literate Society: The Report of the Committee on Reading of the National Academy of Education with a Series of Papers Commissioned by the Committee*, 1975). John Carroll, polymath and professor at the Harvard Graduate School of Education, was my advisor when I was a doctoral student there, but I was puzzled about why they would invite *me* (young and in Pittsburgh, far away, I thought, from Cambridge and New York, the centers of my scholarly experience) to write a paper on literacy development. I also did not exactly understand what an invited paper was. However, I also sensed that this invitation was an honor, and maybe even a “career-maker.”

So I called John Carroll to thank him, and also to ask whether it would be appropriate for me to write an essay—a “think piece,” I might have said—with a colleague who was not a scholar but a highly regarded leader in the Pittsburgh public schools, Betty Robinson. She was one of a very small group of African Americans in leadership positions in public schools at the time. He said, “Sure, work with whomever you would like.” He also said that if I wished, he would take a look at it before I formally submitted. He was, of course, offering to guide me in my first try at authoring in the realm of edited volumes.

The paper, “Motivational Aspects of the Literacy Problem,” is of interest here mainly because it represents a marker of both my own and the NAEd’s desire for collaboration between researchers and thoughtful practitioners. In fact, “teachers” were the first audience we named in the abstract: “This paper will be of interest to teachers, school supervisors, educational researchers and curriculum designers, and others interested in the psychology of instruction.”

Since then, the NAEd has been a good home for someone like me who crosses boundaries between practice, policy, and scholarship. The NAEd has been the only place where my dual identity as a scholar and an activist is (more or less) equally valued. With that first invited paper, I thought I was entering a world of collaboration between research and practice or policy. However, I discovered that that world could not be taken for granted. Our “communities of practice” (a term that migrated across the research/practice boundary) have been more difficult to nurture and sustain than I (and many others, I think) expected. Rather than combining scholarship with policy and/or practice, we have had to move back and forth between them. There have been some successes, but our Academy has not become the routine source of vital information and guidance for policymakers, activists, and various kinds of practitioners that we hoped it would be. Why not?

Reflection (appropriate, I hope, for a collection of *memoirs*) leads me

to notice three fundamental tensions between the worlds of scholarly research and those of education policy and practice. These relate to *time scale*, *evidence standards*, and *voicing*.

Time scale. Scholarly research often trails the needs of practitioners and policymakers. It can take a year or more to develop a research plan and find the necessary funding to carry it out. The research may take several years. Then papers and books are written. Scholarly reviews of reports call for editorial changes that take more time. Many important issues (e.g., class size, or the use of tests to define curriculum and promotion practices) can take years of research to provide convincing evidence for or against favored practices. Often, the policy question has faded away before scholars are willing to take a stand on “what research says.” It is easy to understand why policymakers often reach conclusions without consulting our research.

Evidence standards. As scholars, we are trained to distance ourselves from our hypotheses and beliefs. We take the long view, unlike reporters who work to deadlines. We “hedge” our findings. Finally, we call for *more research* (it is hard to imagine a published paper by a graduate student that does not voice this call—although language differs across discipline). Journalists tasked with summarizing research for a general audience may eliminate detail. As a result, scholars often believe that important points have been glossed over.

Voicing. Conducting further research and mining it for relevant findings takes time, of course. However it is scholars’ insistence on considering the pros and cons of all possible options that often leads policymakers to give research short shrift. In more than one political “back room,” I have been asked to please get my scholarly colleagues to tell them what we *do* know. Policymakers may not like our conclusions, but most will make a thoughtful effort to inform their constituencies and develop action plans that have a chance of working. Like policymakers, teachers want to understand our arguments, but they must take them even further. They need usable tools to carry into the classroom.

In the more recent past, I brought the NAEd an opportunity to develop policy statements as a new U.S. president took office. Channels were open, and the administration wanted the best knowledge scholars had produced to inform the education plan. The NAEd agreed right away. Within the 6 months before the president had to release the plan, we believed we could identify the most pressing issues, assemble working groups, and issue reports that included recommendations. Yes, it would be difficult, but no one thought the idea was crazy (or at least I did not, and my colleagues did not want to say “no” to this project). To save time, we hired a professional writer (a journalist) who would attend the meetings and rewrite the committees’ drafts. In addition, in the interest of

time, the scholars named to the working groups agreed to relinquish the demand for proof and certainty that tends to cause delay.

Some of the groups worked well together and were able to produce reports in several months. In other cases, however, the scholars (who had been deliberately chosen for their divergent views) could not agree. Given time, they might have resolved their differences or decided to issue a contrary opinion, but we did not have time. The problem of voicing also arose. One committee chair could not accept the journalist's drafts. The journalist was a skilled education writer familiar with simplifying complex information, but for one highly technical topic lacked the necessary expertise.

In the end, the reports were released too late to play a visible role in policy formulation. The work met scholarly standards, but the moment of opportunity for political influence had passed. Reflecting on this now, it strikes me that the NAEd has made some progress in taking up important issues at a pace that is comfortable for scholars but still reaches a broad audience. It is still the case, however, that many of the people we wish we could influence do not even know that the NAEd exists. However, maybe I should stop worrying. We have had great success in incorporating younger scholars into our conversations through our two fellowship programs. Perhaps that is our most important role—and perhaps answers to dilemmas such as the ones I have raised here will come from them.

Education in the Courtroom: The Need for Unbiased Research

*David S. Tatel*¹

The National Academy of Education's 50th anniversary offers a perfect opportunity to consider the role of education research in policy-making and judicial decisions. When I chaired the board of the Spencer Foundation, which was led by presidents Lawrence Cremin and then Patricia Graham as well as an extraordinary board, I saw superb, unbiased scholars drawing objective conclusions from methodologically sound research—just what policymakers and courts need to resolve important issues involving education.

However, during my years as a civil rights and education lawyer and as a federal judge, I have seen too many scholars replace objective research with ideological advocacy. By this I mean that many scholars have become what David Cohen refers to as “partisans.” Writing op-ed columns and debating each other at conferences and talk shows, these partisans conclude with absolute certainty—and often with dismissive scorn for opposing views—that their studies, often suffering from serious methodological flaws, vindicate their own views of vouchers, teacher quality, desegregation, education funding, affirmative action, or whatever the issue may be. These partisans often appear in our courts as signatories to amicus briefs and hired expert witnesses.

Certainly, scholars have a critical role to play in debates about education policies, and they can help courts resolve legal issues involving edu-

¹ David S. Tatel is a Judge on the United States Court of Appeals for the District of Columbia Circuit. He was elected to the National Academy of Education in 2003.

cation. In my own and other judges' experience, however, many education researchers have crossed the line that distinguishes objective analysis from partisan advocacy. It is this transformation from scholar to advocate that deprives policymakers and the courts of the objective research and expert judgments they need to make wise decisions.

National Academy of Education Members

Current Members¹

Bruce M. Alberts
Anthony J. Alvarado
James D. Anderson
Richard C. Anderson
Alexander W. Astin
Richard C. Atkinson
Hiroshi Azuma
Thomas R. Bailey
Bernard Bailyn
Eva L. Baker
Deborah Loewenberg Ball
James A. Banks
W. Steven Barnett
Hyman Bass
Isabel L. Beck
Miriam Ben-Peretz
Carl Bereiter
David C. Berliner
Charles E. Bidwell
Paul Black
Derek Bok

Hilda Borko
John Brademas
John D. Bransford
Jeanne Brooks-Gunn
John Seely Brown
Jerome S. Bruner
Anthony S. Bryk
Eamonn Callan
Martin Carnoy
Prudence Carter
Courtney B. Cazden
Stephen Ceci
P. Lindsay Chase-Lansdale
Micheline T.H. Chi
Charles Clotfelter
Paul Cobb
Marilyn Cochran-Smith
David K. Cohen
Michael Cole
Allan Collins
James P. Comer

¹ The National Academy of Education members who are regular members, members emeriti, foreign associates, and foreign associates emeriti.

* Served as president of the National Academy of Education.

Lambros Comitas
K. Patricia Cross
Mihaly Csikszentmihalyi
Larry Cuban
William Damon
Linda Darling-Hammond
Erik De Corte
Andrea A. diSessa
Robert Dreeben
Greg J. Duncan
Jacquelynne S. Eccles
Kieran Egan
Ronald G. Ehrenberg
Margaret A. Eisenhart
Richard F. Elmore
Edgar Epps
Frederick Erickson
Elizabeth Fennema
Michael J. Feuer*
Robert E. Floden
Sarah Warshauer Freedman
Susan H. Fuhrman*
Michael Fullan
Adam Gamoran
Patricia Gándara
David Pierpont Gardner
Howard Gardner
James Paul Gee
Carol Gilligan
Herbert P. Ginsburg
Gene V. Glass
Nathan Glazer
Susan R. Goldman
Louis Gomez
Thomas L. Good
Edmund W. Gordon
Patricia Albjerg Graham*
Sandra Graham
Hanna Holborn Gray
James G. Greeno
Pamela Grossman
John T. Guthrie
Kris D. Gutiérrez

Amy Gutmann
Edward H. Haertel
Kenji Hakuta
Eric A. Hanushek
John F. C. Harrison
Robert M. Hauser
Shirley Brice Heath
James J. Heckman
Larry V. Hedges
Jeffrey Henig
Paul W. Holland
Jacqueline Jordan Irvine
H. Thomas James
Jack Jennings
Susan Moore Johnson
Carl F. Kaestle*
Sharon Lynn Kagan
David Kaplan
James A. Kelly
Jeremy Kilpatrick
Walter Kintsch
David Kirp
Michael Kirst
David Klahr
Daniel M. Koretz
Diana Slaughter Kotzin
Helen F. Ladd
Gloria Ladson-Billings
Ellen Condcliffe Lagemann*
Magdalene Lampert
Judith E. Lanier
Sara Lawrence-Lightfoot
Marvin Lazerson
Carol D. Lee
Valerie E. Lee
Richard Lehrer
Hope Jensen Leichter
Henry M. Levin
Robert A. LeVine
Richard J. Light
Marcia C. Linn
Robert L. Linn
Judith Warren Little

Susanna Loeb
Dan C. Lortie
George F. Madaus
James G. March
Kathleen McCartney
Lorraine M. McDonnell
Wilbert J. McKeachie
Milbrey Wallin McLaughlin
Michael S. McPherson
Douglas L. Medin
Hugh Mehan
Deborah W. Meier
John W. Meyer
Jeffrey Mirel
Robert J. Mislevy
Elizabeth Moje
Luis C. Moll
Robert P. Moses
Pamela A. Moss
Richard J. Murnane
Guy R. Neave
Anna Neumann
Sonia Nieto
Harold J. Noah
Nel Noddings*
Pedro Noguera
Jeannie Oakes
Michael A. Olivas
Ingram Olkin
David R. Olson
Gary Orfield
Fritz K. Oser
Annemarie Sullivan Palincsar
Roy D. Pea
P. David Pearson
James W. Pellegrino
David N. Perkins
Paul E. Peterson
Penelope L. Peterson
Denis C. Phillips
Andrew C. Porter
Alejandro Portes
Sophia Rabe-Hesketh

Stephen W. Raudenbush
Diane Ravitch
Sean Reardon
William J. Reese
Lauren B. Resnick
Barbara Rogoff
Thomas A. Romberg
Mike Rose
Sheldon Rothblatt
Cecilia Rouse
Brian Rowan
Robert Rueda
Rubén G. Rumbaut
Michael Rutter
Gavriel Salomon
Manabu Sato
Geoffrey B. Saxe
Marlene Scardamalia
Leona Schauble
William H. Schmidt
Barbara Schneider
Alan H. Schoenfeld
Robert Schwartz
Donna E. Shalala
Richard J. Shavelson
Lorrie A. Shepard*
Lee S. Shulman*
Robert S. Siegler
Judith D. Singer
Robert Slavin
Marshall S. Smith
Catherine E. Snow
Margaret Beale Spencer
James Spillane
Claude M. Steele
Robert J. Sternberg
James W. Stigler
Deborah J. Stipek
Sidney Strauss
Kenneth A. Strike
Marcelo M. Suárez-Orozco
David S. Tatel
Marta Tienda

William G. Tierney
Judith Torney-Purta
David Tyack
Guadalupe Valdés
Deborah Lowe Vandell
Maris A. Vinovskis
Noreen Webb
Bernard Weiner
Lois Weis
Roger P. Weissberg
Finis Welch
Amy Stuart Wells

Clifton R. Wharton, Jr.
Carl E. Wieman
John B. Willett
John Willinsky
J. Douglas Willms
Mark R. Wilson
Suzanne Wilson
William Julius Wilson
Sam Wineburg
Hirokazu Yoshikawa
Kenneth Zeichner

In Memoriam¹

James Allen
Gregory Anrig
Eric Ashby
Stephen K. Bailey*
C. Edward Beeby
Joseph Ben-David
Bruno Bettelheim
Benjamin S. Bloom
Howard R. Bowen
Ernest L. Boyer
Urie Bronfenbrenner
Jere Brophy
Harry S. Broudy
Ann L. Brown*
William Brownell
Guy Buswell
Roald F. Campbell
John B. Carroll
Robbie Case
Jeanne S. Chall
Francis Chase

John L. Childs
Burton B. Clark
Kenneth R. Clark
James S. Coleman
George S. Counts
Lawrence A. Cremin*
Alonzo Crim
Lee J. Cronbach
Michel Crozier
Allison Davis
Vasilii Davydov
Elliot W. Eisner
Erik H. Erikson
John H. Fischer
John C. Flanagan
William Frankena
Anna Freud
Nathaniel Gage
Robert M. Gagne
John W. Gardner
Arthur I. Gates

¹ The National Academy of Education members and foreign associates who are now deceased.

*Served as president of the National Academy of Education.

Jacob W. Getzels
Eleanor Gibson
Robert Glaser*
John I. Goodlad
Nelson Goodman
Thomas F. Green
Maxine Greene
Elizabeth P. Hagen
Maureen Hallinan
A. H. Halsey
Oscar Handlin
Fred H. Harbison
Kjell Härnqvist
Giyoo Hatano
Robert Havighurst
Alexander Heard
Jurgen Herbst
Father Theodore M. Hesburgh
Roger Heyns
Ernest Hilgard
George Hillocks, Jr.
Richard Hofstadter
Sidney Hook
Cyril O. Houle
Harold Howe II
Everett Hughes
Torsten Husén
Bärbel Inhelder
Philip W. Jackson
Michael B. Katz
Francis Keppel
Alan C. Kerckhoff
Clark Kerr
Kazuyuki Kitamura
Wallace E. Lambert
Paul Lazarsfeld
Seymour M. Lipset
Alexander Luria
Fritz Machlup
Jacques Maritain
Mark A. May
Thomas R. McConnell
Margaret Mead
Robert Merton

Samuel Messick
Martin Meyerson
Jacob Mincer
Frederick Mosteller
Michio Nagai
Ewald Nyquist
John Ogbu
Frederick Olafson
James Perkins
Richard Peters
Philip Phenix
Jean Piaget
Sidney L. Pressey
David Riesman
Lord Robbins
Nathan Rotenstreich
Mary B. Rowe
Frederick Rudolph
Seymour B. Sarason
Israel Scheffler
Wilbur Schramm
Theodore Schultz
Joseph J. Schwab
Albert Shanker
Brian Simon
Theodore R.Sizer
Bunnie O. Smith
Richard Snow
George Spindler
Julian C. Stanley
Harold W. Stevenson
Patrick Suppes*
Jan Szczepanski
William L. Taylor
Robert L. Thorndike
Martin Trow
Henry Trueba
Ralph W. Tyler*
Robert Ulich
Philip E. Vernon
Helen M. Walker
Franz E. Weinert
Willard Wirtz