bsp

volume 2 issue 1
2016

spotlight
pre-kindergarten interventions

spotlight editor Ron Haskins

featured topic
health

table of contents

volume 2 issue 1
2016

editors' note

Welcome to the third issue of *Behavioral Science & Policy* (*BSP*). This issue includes a Spotlight Forum on prekindergarten (pre-K) education programs, an article on improving kindergarten through 12th grade school performance, and articles on health-related beliefs and habits.

Given that *BSP* is still in its infancy, it is perhaps appropriate to focus this issue's Spotlight on early childhood education and development programs. **Ron Haskins**, of the Brookings Institution, served as guest editor of this Spotlight section. He commissioned a terrific cluster of articles that examine the effectiveness of existing American early childhood interventions, including Head Start, state pre-K programs, and home visiting programs. These articles examine the extent to which pre-K interventions foster positive development and improve school readiness among children of low-income families, and they suggest ways in which these programs might be enhanced and better coordinated. Collectively, the spotlight articles provide a valuable, evidence-based road map for maximizing the potential of early childhood interventions. In particular, they call for improved theoretical conceptualization of how and why early childhood programs succeed, which is critical for driving programmatic innovations and enhancing methodology for measuring program effectiveness.

Picking up after pre-K, an open contribution to this issue explores how policymakers might improve school performance in kindergarten through grade 12, drawing on insights from education and social psychology research. **Brian Gill**, **Jennifer Lerner**, and **Paul Meosky** observe that most school systems rely on an outcome-based form of accountability (notably, high-stakes testing), with mixed results. The authors identify three additional forms of accountability (rule based, market based, and professional) and review the evidence for their effectiveness in promoting school improvements. Ultimately, the authors recommend multiple complementary approaches. They propose that professional accountability shows the greatest promise for further development—for example, through greater classroom transparency, peer observation, coaching, and advanced teacher certification. This article is timely for American schools, as last year's passage

of the Every Student Succeeds Act by the U.S. Congress allows states greater latitude in designing their accountability policies.

Finally, we include two articles that emerged from a 2015 workshop on health and well-being that was sponsored by the University of Southern California and cohosted by the Behavioral Science & Policy Association (BSPA). These articles highlight the importance of unconscious mental processes in driving health beliefs and behaviors. First, **Wendy Wood** and **David Neal** observe that although public health interventions have succeeded in increasing people's knowledge of healthy behaviors and their intentions to engage in them, these programs typically fail to achieve sustained action by the public. The authors draw on behavioral research to identify keys to healthy habit formation that have produced sustained improvements in eating and exercising behaviors, especially when implemented together: repetition of desired behaviors, stable contextual cues to trigger habit performance, and intermittent rewards to reinforce these habits. The authors also provide keys to neutralizing unwanted habits. In a second article, **Norbert Schwarz, Eryn Newman, and William Leach** review how ease of cognitive processing influences the believability of facts and myths, such as the claimed link between vaccines and autism. They draw on these insights to suggest a number of policy approaches for making truthful public health information stick and harmful health myths fade.

As always, we invite readers to contribute feedback to our editorial office so that we can improve future issues of *BSP.* We also encourage readers to engage with our parent organization, BSPA. This rapidly growing community consists of behavioral scientists, policy professionals, and other practitioners who have a shared interest in the thoughtful application of rigorous behavioral science research to policy and practice in ways that serve the public interest. In addition to joining BSPA, readers can sign up to receive our weekly and monthly roundups summarizing the latest news and developments in behavioral policy, and check out our PolicyShop blog that explores topics in greater depth. Readers can also engage with our community by attending a BSPA workshop or annual conference. Finally, we encourage you to volunteer time or resources to BSPA, which is a nonprofit, public benefit organization. Information is available on our website, http://www.behavioralpolicy.org.

We look forward to continuing to bring readers articles that pass our unique dual peer-review (involving both disciplinary and policy editors) and are professionally edited to enhance applicability and accessibility to a broad and diverse audience. Our next issue will feature reports from several BSPA working groups that identify promising approaches for new behavioral policy interventions at the federal level. We will also present a fascinating

essay by Jason Furman (chair of President Obama's Council of Economic Advisors) in which he describes how behavioral insights can support macroeconomic policy. And, as usual, we'll present the latest empirical findings, essays, and reviews from leading researchers in the behavioral policy community.

Craig R. Fox & Sim B Sitkin
Founding Co-Editors

American policy on early childhood education & development:
Many programs, great hopes, modest impacts

Ron Haskins

Spotlight Introduction Summary.   The primary motivation for this Spotlight section on early childhood programs is to assess whether and to what degree they are successful in promoting the development and school readiness of children from poor families. Conflicting claims abound over the effectiveness of public programs such as Head Start and state-funded prekindergarten (pre-K) and whether they are meeting the intended goals of preparing disadvantaged children for school and boosting the overall development of served children and their families. The disappointing results of the federal study of Head Start (the Head Start Impact Study [HSIS], reported in 2010) showing that the immediate positive impacts on children's achievement quickly faded[1] added fuel to the evolving debate on what does and does not work in publicly funded early childhood education. Because other pre-K studies, conducted over similar time periods as the HSIS, have demonstrated more promising results, the hope remains that these programs can significantly boost children's development and school readiness. High-quality evaluations of state pre-K programs show that some produce substantial gains in intellectual development,[2] yet many programs do not. In addition, few of these studies have shown long-term impacts on children. Another popular approach to advancing family and childhood development is home visiting programs (HVPs). Trained professionals or paraprofessionals work with new mothers, improving their child-rearing skills and assisting with life issues such as perinatal depression and employment. As with Head Start and state pre-K programs, the benefits of HVPs are often modest or overstated. There are also many unresolved issues about both the long-term impacts of these programs and the nagging

but pressing question of whether successful interventions can produce good results when implemented at hundreds or even thousands of sites nationwide. Clearly, improvements are needed in setting early education and development policy and in advancing the research that will point the way forward. The articles in this Spotlight address these and other issues faced by Head Start, state pre-K, and HVPs and offer a host of solutions for educational policymakers to consider.

This Spotlight feature includes four articles on these three large-scale early childhood programs and an article that proposes better coordination and improvements in these programs in order to achieve maximum impact. A major goal of these programs is to help close the achievement gap between poor children and their more advantaged peers. The evidence that children from poor families lag far behind children from wealthier families when they enter the public schools is overwhelming. Educational disadvantage, one of the key causes of high poverty rates and stagnant economic mobility, begins during the earliest years of life and is well established by the time children enter public schools. Worse, disadvantaged children rarely catch up.[3,4] Few argue with the laudable goal of leveling the educational playing field, yet how best to do so is far from established, and there is disagreement among the experts on what the public can expect from early childhood programs. It is widely believed that high-quality early childhood programs are a major strategy that can help equalize life outcomes for poor children and minorities. But as this Spotlight section will show, many complexities are involved in conducting and organizing early childhood programs, and only high-quality programs produce significant impacts.

The Birth of Head Start and U.S. Early Education Policy

Research on human development using gold-standard random-assignment designs provides compelling evidence that early life experiences can be manipulated to enhance development.[5] Further, two immensely successful early childhood programs initiated in the 1960s and 1970s, the Perry Preschool Project in Michigan[6] and the Abecedarian Project in North Carolina, stimulated great hope that early childhood programs could substantially offset the effects of poverty.[7] (Farran

and Lipsey limn these two experimental programs in greater detail in their Spotlight article.) The architects of the Perry Preschool and Abecedarian Projects both accomplished the remarkable feat of following children into their 40s and found that those who had participated in either early childhood program continued to excel during adolescence and adulthood. In one or both projects, compared with controls, children enrolled in the intervention programs were less likely to be in special education, be retained in grade, be arrested, have a teen birth, or go on welfare; they were also more likely to graduate high school, attend and graduate college, and be employed.

The Head Start program was the first and is still the most notable federal effort to enter the preschool arena. Sargent Shriver, a Kennedy family member and the head of President Johnson's War on Poverty, had visited a preschool program in Nashville that was similar to the Perry Preschool Program; he quickly formed the view that preschool should be a major weapon in the War on Poverty. Shriver then convinced President Johnson of the importance of early childhood intervention programs. In short order, the Johnson administration sponsored legislation that included funds for the new program. As a result, about 500,000 poor children participated in Head Start's inaugural program in the summer of 1965.[8,9]

With Head Start as its anchor, the federal government fitfully began to expand its commitment to early childhood programs. In 1969, President Nixon reflected the growing bipartisan view that early childhood was an especially important period of human development when he told Congress that he was committed to helping children during their crucial first 5 years of life. Soon John Brademas (D-IN) in the House and Walter Mondale (D-MN) in the Senate took Nixon up on his support for early childhood programs and introduced legislation that would have eventually provided $5 billion

(about $32 billion in today's dollars) for early childhood programs designed and conducted by local communities. The bill handily passed both houses of Congress before being vetoed by Nixon, primarily on the grounds that the nation should not support "communal" forms of child rearing.[10]

The death of the Brademas–Mondale bill put a temporary chill on federal involvement in early childhood programs. Yet as federal social policy, programs, and spending expanded dramatically over the next four decades, early childhood programs, aimed at both boosting the development of poor children and providing routine care for children while parents worked or attended school, grew almost continuously.

Four Main Categories of Early Childhood Programs: Head Start, State Pre-K, Home Visiting Programs, and Day Care

Federal and state investments in preschool now total around $34 billion annually; this high spending level demonstrates the magnitude of the nation's buy-in to the theory that public support for children during their early years is vital and that poor children's participation in high-quality programs during these years can help get them off to a good start. Table 1 provides an overview of spending on the major federal and state early childhood education programs.

To ground readers in each of the four main early childhood programs, I offer a brief review of the major characteristics of each type of program and the way in which each is examined in the Spotlight articles in this issue.

Head Start

Although Head Start was born as a comprehensive preschool program—with goals that include social and intellectual development, nutrition assistance, and health management—since its inception, the nation has adopted numerous other children's health and nutrition programs. As a result, the need for a comprehensive preschool program is not as great as it was when Head Start began more than a half century ago.

The Spotlight article that offers policy guidance on how to reform Head Start to keep pace with the times, authored by Sara Mead and Ashley LiBetti Mitchel of Bellwether Education Partners, underscores this point

Table 1. Summary of Spending on Major Early Childhood Programs, 2015 ($ billions)

Program	Spending[a]
Federal	
Head Start & Early Head Start	8.6
Child Care Development Block Grant	5.3
Child Care Food Program	3.1
Child and Dependent Care Tax Credit (Tax code)	4.5
Dependent Care Assistance Program (Tax code)	0.9
Individuals with Disabilities Act (IDEA)	0.8
Home Visiting	0.4
Preschool Development Grants	0.25
Temporary Assistance for Needy Families (TANF) Block Grant	1.2[b]
State	
State Preschool	6.2[c]
TANF Maintenance of Effort (MOE)	2.5[d]
Total	**33.9**

a. Data consist of updated figures for FY 2016 from *Congressional Research Service* report titled "Early Childhood Care and Education Programs: Background and Funding" by Karen E. Lynch. Note that spending figures do not include Title XX, Grants (Social Services Block Grant), or Education for the Disadvantaged- Early Childhood Block Grants: Title 1, Part A (more information available at http://pennyhill.com/jmsfileseller/docs/R40212.pdf).

b. Latest available data for 2014; assume constant expenditure level. Figure drawn from "TANF Spending on Child Care Up Slightly in 2014" available at http://www.clasp.org/issues/child-care-and-early-education/in-focus/tanf-spending-on-child-care-up-slightly-in-2014/.

c. Report available at http://nieer.org/sites/nieer/files/2015%20Yearbook.pdf. Figure includes federal TANF funds directed toward preschool at states' discretion. In 2014–2015 Indiana began offering a state-funded pre-K program with $1 million in state funding. Because it served less than 1% of 4-year-olds, these funds are not reflected in the funding total.

d. Latest available data for 2014; assume constant expenditure level. Figure drawn from "TANF Spending on Child Care Up Slightly in 2014" available at http://www.clasp.org/issues/child-care-and-early-education/in-focus/tanf-spending-on-child-care-up-slightly-in-2014/.

and calls for a greater focus on the program's primary goal of enhancing kindergarten readiness by stimulating the intellectual and socioemotional development of enrolled children. The authors argue that Head Start must continue to evolve in this focus if it is to remain relevant in the face of massive upscaling of state pre-K programs that more narrowly target school readiness. Mead and Mitchel aptly emphasize that Head Start should reduce its overreliance on regulation monitoring as the primary means of program control and increase

the use of outcome measures as a performance barometer. They also outline important ways in which Head Start can be improved through better coordination with the other early childhood programs, triage of services based on need, and tweaks to the designation renewal system that will increase the number of quality Head Start providers in the pipeline to replace failed programs. An important element of Mead and Mitchel's discussion of Head Start is a review of recent reform efforts initiated by the Obama administration, which they believe hold promise for improving Head Start outcomes, yet they also lament that the current statute prevents the administration from acting more aggressively both to reduce the number of noneducational services programs are required to provide and to increase grantees' flexibility to focus services on children's and communities' actual needs.

State Pre-K

About 1.38 million children are now enrolled in another important early childhood program, state pre-K. This is more than the approximately 950,000 students enrolled in Head Start. Very few states had pre-K programs until the 1980s, when 23 states initiated them.[11,12] Currently, 42 states and the District of Columbia conduct public pre-K programs.[11] Clearly, many state policymakers felt the need to supplement Head Start, either because it did not enroll enough of the poor 3- and 4-year-olds who resided in the state or because so many children from poor families continued to appear at the schoolhouse door unprepared.

This Spotlight includes two articles on pre-K research because it is necessary to have both the optimists and the skeptics present their cases about whether the evidence from evaluation studies shows that pre-K programs are having positive, lasting effects on the development and school readiness of poor children.

The article by Christina Weiland of the University of Michigan conveys the clear message that high-quality pre-K programs are effective and that research is showing how the programs can be made even more effective. She points out that, unfortunately, all too often policymakers and educators are not using the best evidence-based curricula in the classroom; she underscores this point by noting that one of the most commonly used curriculum in Head Start and state pre-K programs received an effectiveness rating of zero

from the What Works Clearinghouse, an arm of the U.S. Department of Education. She also makes a strong case for enhanced evaluation and development of domain-specific curricula (for example, in reading and math) over more standard comprehensive, whole-child curricula. Weiland examines in detail what constitutes quality in preschool education as well as the role of teachers, training, coaching, and curriculum in achieving quality. She then summarizes the results of an evaluation study she and her colleague Hirokazu Yoshikawa of New York University conducted of a high-quality pre-K program in the Boston public schools. Weiland shows that the Boston program, which involved 67 schools and over 2,000 students, had major positive effects on children's literacy, language, and mathematics skills. Given the size of the study and the impressive magnitude of the results, she concludes that major impacts on children's intellectual development are possible even in a large-scale program.[13]

The second state pre-K–focused Spotlight article, authored by Dale Farran and Mark Lipsey of Vanderbilt University, provides a more skeptical take on the evidence of benefit to children. They begin their review by lamenting that there is no common definition of what constitutes a pre-K program. Rather, the 40-plus programs run by states vary greatly in student-to-teacher ratios, teacher training, curricula, program goals, hours of operation, and many other characteristics. Therefore, it is almost senseless to categorize the benefits of "state pre-K" without a more distinct set of common characteristics and practices. Moreover, they argue that most of the studies on which claims about the effects of state pre-K are based, suffer from serious methodological shortcomings, especially those that examine sustained effects. It is notable that there is only one well-controlled evaluation of the sustained effects of a state pre-K program, which Farran and Lipsey conducted. Their evaluation of the Tennessee Voluntary Preschool Program found positive but modest impacts on measures of early achievement and teacher ratings of preparedness for school at kindergarten entry. But those effects were not sustained past the end of kindergarten and, remarkably, by the second and third grades, children in the control group, who had not attended pre-K, actually scored higher on some achievement measures than did children who had attended pre-K. Farran and Lipsey fairly point out that although state pre-K programs vary greatly, they are common in a

singular trait: none are required to implement program components of models shown to provide long-term benefits to children. And perhaps most bravely, they question whether pre-K programs produce effects on children's development that last beyond the end of the program, implying that by the end of the kindergarten or first grade years of schooling, children who attended pre-K do not perform better than they would have if they had not attended pre-K.

It is difficult to reconcile the Weiland and the Farran and Lipsey reviews. However, the Boston pre-K program that is the major source of Weiland's claims of substantial impacts is not included in the Farran and Lipsey review. The Boston study is not based on a random assignment design, but the design it did use—called *regression discontinuity*—is nonetheless widely considered only a modest step below random assignment.[14] The study is also large in scale and was the result of several years of innovation and improvement, as administrators adopted evidence-based domain-specific reading and math curriculums and built in extensive training and coaching of teachers. One limitation of the Boston program is that the pre-K teachers were paid on the same scale as public school teachers, a rare occurrence in state pre-K programs that renders the Boston program, at $15,000 per student, too pricey for most states. In addition, its regression discontinuity design hampers long-term follow-up, so one of the big questions about pre-K programs—whether they produce long-term effects—will not be answered with the methodological rigor of the initial impacts. However, Weiland and her colleagues are now conducting a large-scale, random-assignment study of the Boston program that follows participating children through the third grade; they will soon be reporting these results. This new study will provide a reliable test of whether the Boston program is producing effects that last several years beyond completion of the pre-K program. Meanwhile, observers looking to cite a rigorous study that shows big short-term impacts should look to the Boston evaluation study that has already been published.[13]

Day Care

A word is in order about federal and state day care programs, funded primarily by the Child Care and Development Block Grant (CCDBG) with around

$5.3 billion of federal and state funds (see Table 1). Lots of rhetoric about quality programs surrounds the discussion of the CCDBG, as suggested by the term *Child Development* in the program's title, but the facilities funded by the program show an enormous range of quality. A few of the facilities are of high quality and probably do promote child development, but most of the facilities are of mediocre quality or worse. It seems doubtful that many of these facilities actually promote child development, and some may even impede it.[15,16] Because the major goal of this program is to provide safe child care and not developmental care, a separate article on the CCDBG is not included in this issue. But readers should be aware that a majority of children, especially poor children, are enrolled in day care facilities that do not promote their development or prepare them for school.

Home Visiting Programs

Most preschool programs make at least some attempt to involve parents because they are so central to their children's development.[17] But rather than just involving parents, HVPs focus specifically on helpings parents, especially mothers, improve their child-rearing skills. Most of these programs send a trained home visitor into the child's home to routinely meet with the mother and child, sometimes beginning during the prenatal period, and lasting for a year or two. The roots of home visiting as an intervention date back at least to Florence Nightingale (1820–1910) and her emphasis on both health issues and home issues of safety and infant development among poor mothers.[18] A variety of rigorously evaluated model HVPs exist. Most follow a set of activities that the home visitor uses to help teach mothers how to engage in productive activities with their children. The general goal is to get mothers to be verbally responsive to their infants and young children and to respond to their children's signals. The programs also help mothers resolve personal issues, with services such as treatment referrals for depression, employment guidance, and training program placements.

Until recently, most HVPs were initiated by and paid for primarily with state funds (although the federally funded Early Head Start is an exception). Then, in the Patient Protection and Affordable Care Act (better known as Obamacare) enacted in 2010, a federal

program was created that provided $1.5 billion over 4 years for states to expand home visiting, primarily through the use of model HVPs that show strong evidence of having positive effects on mothers or children.[19] The U.S. Department of Health and Human Services worked with the Mathematica Policy Research firm to conduct a systematic review of research on HVPs to determine which model programs had strong evidence of impacts on important outcomes such as improved child health, reduced incidence of child abuse and neglect, and improved maternal health. The home visiting funds were funneled through states with the caveat that states had to spend 75% of the funds on one or more of the 11 model programs identified at that time by the U.S. Department of Health and Human Services as being evidence based.[20] The federally and state-funded HVPs currently serve more than 115,000 parents and children in 787 counties throughout the nation at a cost of about $400 million a year, and several new evidence-based model programs have been approved for use.

The Spotlight review article on home visiting was written by Cynthia Osborne of the Lyndon B. Johnson School of Public Affairs at the University of Texas, who is the chief evaluator for the State of Texas's home visiting program, the largest in the nation. Osborne recognizes the importance of the federal requirement that the majority of federal HVP funds go to evidence-based programs. However, she stipulates that several additional points need to be addressed if HVPs are to achieve maximum results. Perhaps most notably, she calls for better matching of the particulars of an HVP model to the specific needs of a family and for model developers to identify the specific aspects of their model programs that are the crucial elements for producing specific outcomes. (This point is reminiscent of Farran and Lipsey's emphasis on the importance of specifying the features of a curriculum that are the active ingredients in stimulating a child's development). In this way, communities can choose the best HPV model to meet their specific needs, and home visitors can be sure to implement the elements of that model with fidelity. She also notes that continued innovation in HVPs is vital to keep up with the evolving problems experienced by poor mothers, for example, by shifting focus from smoking cessation assistance to weight loss programs for new mothers.

Expansion

These four literature reviews show that early childhood programs are, at the very least, promising. Several individual programs, including the Perry Preschool Program, the Abecedarian Project, the Chicago Child-Parent Centers,[21] the Tulsa pre-K program, and now the Boston pre-K program have produced remarkable and in some cases lasting impacts on children's development. The same is true of the Nurse-Family Partnership home visiting program[22] and perhaps some of the other home visiting programs labeled *evidence based* by the U.S. Department of Health and Human services.[23] But can these programs be coordinated to maximize the impacts they could achieve and move the nation toward a seamless system of early childhood intervention programs? Ajay Chaudry, a former senior official at the U.S. Department of Health and Human Services, and Jane Waldfogel of Columbia University propose reforms that would create a strategy to get the most out of the nation's early childhood programs, especially if the federal government is willing to add the significant additional funds needed to create such a coordinated early childhood system.

The new system they propose has four major features. The first is a greatly expanded parental leave program so that parents would have the opportunity to spend 12–16 weeks with their newborns to establish early parent–child bonding. The second feature is a subsidy for the purchase of regulated child care that would be provided through expansion and reform of the CCDBG for low- and moderate-income families and of the child care tax credit for families with enough earnings to pay federal income taxes. The third and most expansive provision is to create a universal pre-K program for all children beginning at age 3 years. In effect, this recommendation means that public education in the United States would begin at age 3. The pre-K system would be owned and operated by local government, but the federal government would provide matching funds over the first decade of the new system. Finally, Chaudry and Waldfogel's proposed early childhood system would be completed by a "narrowly targeted, intensive, and comprehensive" initiative aimed at infants and children in families who live in deep poverty or who have serious developmental problems. This new initiative would meld Head Start, Early Head

Start, and HVPs into a coherent system with centers and home visitors located in the nation's most disadvantaged communities.

Conclusion

The nation's early childhood landscape includes a growing home-visiting movement, an even bigger state pre-K movement, a venerable Head Start program, and a very big day care sector that serves upward of 2.2 million kids each month.[24] Given that the nation is counting on this array of programs to be a leading weapon to reduce poverty and promote economic mobility, we need to carefully assess how these programs are working in order to make better decisions about whether federal spending should be increased and, if so, what programs are our best bets.

An important part of the context for policymaking on early childhood programs is whether the public supports the programs and is willing to pay the bill. In the case of children from low-income families, all of the early education discussed here is subsidized or completely paid for by tax dollars. Public support is therefore vital. Polls usually show that the public does support these programs. A 2015 poll of 800 registered voters conducted by Public Opinion Strategies and Hart Research, for example, found the 54% of respondents said they would "hold a more positive view" of any presidential candidate who supported improving early education. In addition, when interviewers asked respondents to rank the importance of a list of policy topics, 89% said that children getting a "strong start in life" is "extremely" or "very" important. A previous poll by the same polling companies found that 76% of respondents supported a 2012 proposal by President Obama to spend $100 billion over 10 years to expand preschool programs for low- and moderate-income families.[25] Voters likely favor significant spending on early education because they read numerous stories in the press that preschool is successful, despite the fact that that success may be overstated by both the media and politicians, as several of our Spotlight authors forthrightly point out.

At the risk of incurring the wrath of advocates for these programs and perhaps even some of the authors in this Spotlight feature, the modest conclusion that enjoys the greatest support from high-quality research is that good programs can achieve immediate impacts and some exceptionally high-quality programs can even produce long-term impacts, especially in reducing grade retention and avoiding placement in special education. But, as this Spotlight review seems to make clear, many of the early childhood programs now operating in communities throughout the nation are producing, at best, short-term impacts. The field of early intervention still has a lot to learn, and the jury is out on whether these programs can help the nation reduce poverty and increase economic mobility.

author affiliation

Haskins, Center on Children and Families, Brookings Institution, Washington, DC. Corresponding author's e-mail: rhaskins@brookings.edu

References

1. Puma, M., Bell, S., Cook, R., Heid, C., Broene, P., Jenkins, F., Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study: Final report* (OPRE Report 2012-45). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.
2. Gormley, W. T., Jr., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41,* 872–884.
3. Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap*. Washington, DC: Brookings Institution Press.
4. Lee, V. E., & Burkam, D. T. (2002). *Inequality at the starting gate*. Washington, DC: Economic Policy Institute.
5. Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development*. Washington, DC: National Academies Press.
6. Schweinhart, L. J., Barnett, W. S., & Belfield, C. R. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40*. Ypsilanti, MI: High/Scope Press.
7. Ramey, C. T., Sparling, J. J., & Ramey, S. L. (2012). *Abecedarian: The ideas, the approach, and the findings*. Los Altos, CA: Sociometrics.
8. Zigler, E., & Muenchow, S. (1992). *Head Start: The inside story of America's most successful educational experiment*. New York, NY: Basic Books.
9. Vinovskis, M. S. (2005). *The birth of Head Start: Preschool education policies in the Kennedy and Johnson administrations*. Chicago, IL: University of Chicago Press.
10. Steiner, G. Y. (with Milius, P. H.). (1976). *The children's cause*. Washington, DC: Brookings Institution.
11. Barnett, W. S., Friedman-Krauss, A. H., Gomez, R., Horowitz, M., Weisenfeld, G. G., Brown, K. C., & Squires, J. H. (2016). *The state of preschool 2015: State preschool yearbook*. New Brunswick, NJ: National Institute for Early Education Research.

12. Mitchell, A. (2001). *Kindergarten programs in the states: Trends and issues*. Washington, DC: Institute of Education Sciences, Education Research Information Center.

13. Weiland, C., & Yoshikawa, H. (2013). Impacts of a pre-kindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84,* 2112–2130.

14. Porter, K. E., Reardon, S. F., Unlu, F., Bloom, H. S., & Robinson-Cimpian, J. P. (2014, November). *Estimating causal effects of education interventions using a two-rating regression discontinuity design: Lessons from a simulation study* [Working paper]. Available from http://www.mdrc.org/publication/estimating-causal-effects-education-interventions-using-two-rating-regression

15. Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009.* Washington, DC: U.S. Department of Health and Human Services.

16. Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–31). Baltimore, MD: Brookes.

17. Kalil, A., Ryan, R., & Corey, M. (2012). Diverging destinies: Maternal education and the developmental gradient in time with children. *Demography, 49,* 1361–1383.

18. Wasik, B. H. (1993, Winter). Staffing issues for home visiting programs. *The Future of Children, 3*(3), 140–157.

19. Haskins, R., & Margolis, G. (2015). *Show me the evidence: Obama's fight for rigor and results in social policy.* Washington, DC: Brookings Institution Press.

20. U.S. Department of Health and Human Services. (n.d.). *The maternal, infant, and early childhood home visiting programs: Partnering with parents to help children succeed* [Issue brief]. Retrieved from https://www.acf.hhs.gov/sites/default/files/ecd/home_visiting_issue_brief_2015.pdf

21. Reynolds, A. J. (2000). *Success in early intervention: The Chicago Child-Parent Centers.* Lincoln: University of Nebraska Press.

22. Olds, D., Eckenrode, J., Henderson, C., Jr., Kitzman, H., Powers, J., Cole, R., Luckey, D. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association, 278,* 637–643.

23. U.S. Department of Health and Human Services. (n.d.) *Home visiting evidence of effectiveness.* Retrieved from http://homvee.acf.hhs.gov/models.aspx

24. Chien, N. (2015, November). Estimates of child care eligibility and receipt for fiscal year 2012. *ASPE Issue Brief.* Washington, DC: U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Office of Human Services Policy.

25. Tully, S. (2015, October 20). Poll: Voters increasingly support early education investments. *Edsource Today.* Retrieved from https://edsource.org/2015/poll-voters-increasingly-support-early-education-investments/89281

Evidence for the benefits of state prekindergarten programs: Myth & misrepresentation

Dale C. Farran & Mark W. Lipsey

Summary.   In 2014, New York City launched the historic Pre-K for All program, which massively expanded early education for children in the city. The state followed suit with a budgetary pledge of some $1.5 billion over 5 years to implement full-day universal prekindergarten (pre-K) statewide. Many other states have implemented or expanded state-funded pre-K programs in the last decade, encouraged by claims about expected benefits. However, there is remarkably little scientifically rigorous evidence for these benefits. Claims of pre-K success rely largely on small, experimental, model programs run 50 or more years ago—programs that bear little resemblance to current pre-K implementations. Evidence for program effectiveness is crucial given the current interest in and expansion of state-funded pre-K programs. This review of the evidence raises serious questions about the presumed benefits of state pre-K programs.

The momentum to create or expand state-funded prekindergarten (pre-K) programs increased dramatically in the mid-2000s, encouraged in large part by a perspective article published in *Science* by James J. Heckman that called for investments in early childhood education for disadvantaged children.[1] This charge was quickly adopted by education advocacy groups such as ReadyNation. However, Heckman's conclusions about the benefits of such investments were based on research from the 1960s and 1970s on the effects of two well-funded experimental programs: the Perry Preschool Project and the Abecedarian Project. The Perry Preschool Project was a 2-year intensive intervention that began when children were 3 years old and required substantial parental involvement—researchers visited parents in their homes once a week for an hour and a half. The Abecedarian Project began when children were 6 weeks old and lasted seamlessly until they entered kindergarten. The care covered 8–10 hours a day for 50 weeks of the year. These programs, because of their intense interventions and high cost, have been dubbed "Cadillac programs," and few feel that they are sustainable on a national or state stage.

More recent support for statewide early education programs came in the form of a 2014 White House report titled *The Economics of Early Childhood*

Investments.[2] This report referred to a "deep" research base supporting pre-K and cited the same 1960s and 1970s studies that Heckman referenced. The report went on to state that "dozens of other programs have been rigorously examined since the 1960s" and asserted that high-quality early education will narrow the achievement gap, boost adult earnings, and result in savings of $8.60 for every $1 spent on the program.

Despite various correction efforts by independent watchdogs[3,4]—such as the *Washington Post*'s Fact Checker,[5] which called into question the administration's pre-K cost–benefit claims—belief in the long-term payoff of pre-K programs is so well entrenched in the thinking of policymakers that the state sponsors of a relatively new program, the Great Start Readiness Program in Michigan, were willing to claim long-term, adult benefits even though the program has not existed long enough for such outcomes to be observed.[6] Similarly, in a Texas-commissioned report, researchers Robert Pianta and Catherine Wolcott of the University of Virginia[7] assured policymakers that they need not spend money on Cadillac programs with unsustainable costs because other states have demonstrated that lower cost programs can achieve the same results. The programs they cited do not have adequate research to justify these claims, but they are cheaper.

Claims for Effectiveness Must Align with Reality for Sustained, Appropriate Long-Term Investments in Pre-K Education

The problem that states and society at large face is a genuine one: Children from low-income backgrounds underachieve in school compared with children from higher income families, and this poverty gap now eclipses racial achievement gaps.[8] The presumption that poor children can be made more ready for school by kindergarten entry in a way that will then propel them forward to achievement equal to that of their more privileged peers has been present in this country for 50 years, at least since the creation of Head Start.[9] In this article, we do not question whether this is an important goal; it is.

Our concern is the substantial discrepancy between the actual evidence and the expected benefits of contemporary scaled-up, statewide pre-K programs that have been adopted on the basis of strong claims about how compelling the supporting research is. Indeed,

long-term funding of early education is in considerable danger if the public is overpromised on what it can and should expect from pre-K programs. When a governmental study on the effects of the Head Start early education program was reported in 2010, the results were so dismal that *TIME Magazine* called for an end to Head Start.[10]

In this article, we place the recently reported results[11] (extending to Grade 3) from our study of Tennessee's pre-K program within the spectrum of studies—of varying methodological quality—on similar state-wide programs. First, however, we challenge anyone to define with any specificity what a statewide pre-K program actually is or should be. Next, we consider the outcomes such programs are expected to bring about and examine the research evidence on state pre-K programs. Last, we highlight the difficulties associated with scaling up an effective program even when the intended outcomes are well defined.

Highly Divergent State Pre-K Programs

A quote from a Pew Charitable Trusts analysis sums things up: "While there's a growing consensus on the value of preschool, states disagree on where the programs should be based, who should run them, or how the government should support them."[12] States are doing remarkably different things under the rubric of pre-K. And this situation is continually in flux—many states have recently passed legislation substantially changing or expanding their programs. Nonetheless, most, although not all, state programs more or less follow a public school model in setting parameters, meaning that pre-K classrooms live in public elementary schools, a full day is a standard school day (typically 6 hours), and a full year is a standard 9-month school year.

States differ on whether lead teachers in pre-K classrooms should be licensed and/or have bachelor's degrees. Class size requirements differ as well; for example, New Jersey limits class size to 15 students, but Texas has no specified limit. Most states set the adult-to-child ratio at 1 to 10, but in Texas, it may go as high as 1 to 22, a stark difference.

And whereas most states target their programs to children from low-income families, financial eligibility cutoffs vary widely. For example, many states use the eligibility criteria for the free and reduced-price lunch

(FRPL) programs—135% and 185% of the federal poverty level, respectively. However, North Carolina uses 75% of the state's median income as the cutoff, and Michigan sets it at 250% of the federal poverty level.[13] In the pre-K expansion grants newly funded by the federal government, the eligibility requirement is 200% of the federal poverty level.[14]

However, state pre-K programs are quite similar in one regard: No state requires implementation of a program with components that match the Perry Preschool Project or Abecedarian Project models, even though both have shown long-term benefits for participants. More commonly, programs are required to implement an early childhood curriculum, typically on a state approved list, but those curricula lack distinct evidence of effectiveness and do not share a common vision. In New Jersey, for example, programs are limited to choosing among Tools of the Mind, Creative Curriculum, Curiosity Corner, and HighScope—curricula that differ greatly from each other in content and approach. Research on three of these has found no additional benefits compared with generic practice, according to reviews by the What Works Clearinghouse,[15] and the fourth (HighScope) has not been reviewed (for more information on the What Works Clearinghouse, see the online Supplemental Material).

Very few states invest in monitoring the quality of pre-K programs once funds are awarded to school systems. Monitoring is left up to local school systems. A few states, such as Tennessee and Louisiana, require self-assessments using measurement tools such as the Early Childhood Environmental Rating Scale, but they do not regulate the rigor with which those assessments are done or how the results are used. Only rarely do states require that their public school pre-K programs be rated by the state's Quality Rating and Improvement System (for more on the Quality Rating and Improvement System, see the online Supplemental Material), even though many states do require such an evaluation of community child-care programs serving children of the same age.

Diverging Goals for Pre-K

Not only do pre-K programs differ across states, but also the declared goals of the programs vary—even within a state—and too often those goals are ambiguous. If there were statewide consensus about the goals for pre-K, even with disagreement on the best way to attain them, at least relevant outcomes could be identified and progress toward achieving them could be evaluated. The goals for state pre-K programs are generally more implicit than explicit and include the following variations.

School readiness. Kindergarten readiness frequently appears as a primary objective of public pre-K—especially in the most recent request for proposals for pre-K expansion from the Obama administration.[14] Yet school readiness is often vaguely defined; it usually means some mix of literacy and numeracy skills, proper school behavior, and perhaps motor development skills and health objectives. Strictly speaking, school readiness as a goal includes no assumption of sustained effects beyond kindergarten entry, although the expectation certainly is that reducing the gap at school entry for disadvantaged children will allow them to progress more effectively in later grades.

School achievement. This goal focuses on learned academic content and skills that go beyond simple school readiness and instead are manifest most clearly in performance on achievement tests, grades, and grade retention. The general research literature on the effects of pre-K does not provide much encouragement for the expectation that gains in academic achievement will be sustained for very long. Any beneficial effects have typically been found to diminish well before high school, even for the Perry Preschool Project and the Abecedarian Project.[16] Nonetheless, sponsors and stakeholders, including the White House, expect that state pre-K programs will help close school achievement gaps for minority and economically disadvantaged children in the long run. As such, whatever initial positive effects pre-K may have on achievement are expected by policymakers to carry through into the later grades.

Behavioral outcomes. These are the "noncognitive" outcomes, such as attendance, disciplinary infractions, graduation, employment, and criminal behavior. Many proponents cite beneficial effects on these behaviors as a primary justification for scaling up pre-K programs. These behaviors are not directly reflective of academic achievement but may speak to broader effects on children's persistence and goal orientation. Expectations of such effects are based on the classic

longitudinal studies of the Abecedarian Project and the Perry Preschool Project that found long-term behavioral benefits in some areas, although not consistently on the same outcomes across the two programs. The long-term benefits found in these early programs are also the primary basis for the claims of the cost effectiveness of pre-K.[17]

Outcome Studies: Limited Research and Weak Methods

Relatively few studies of the effects of state-funded, scaled-up pre-K programs have been reported, and they rarely appear in peer-reviewed outlets. We find the most striking features of these studies to be weak methods and limited outcome assessments. Here, we summarize the study designs, outcome variables, and findings from this research—first for effects found at the end of the pre-K year and then for sustained effects through the school years following pre-K.

Short-Term Effects

The methodological rigor of research on the immediate effects of state pre-K programs is not generally strong. Nonetheless, there is an overall pattern of positive effect estimates of sufficient magnitude to suggest likely benefits, particularly on achievement outcomes. Moreover, that conclusion is consistent with the broader research literature on the short-term cognitive effects of early childhood education.[18]

Randomized studies. The most methodologically rigorous design for studying program effects is a randomized control trial (RCT). In an RCT, participants are assigned to treatment and control groups via a chance process that results in no systematic differences between the groups at the beginning of the study. The only RCT of a state pre-K program is the one we have conducted on the Tennessee program, and it is not yet complete.[19] The study was done in 58 school-based pre-K programs that agreed to admit children in random order until the available seats were filled; the remaining applicants served as the control group. For a subsample of one-third of the total sample, outcome data were collected annually to track program effects through the third grade, when the state achievement tests are administered. The outcome measures included

Woodcock-Johnson III achievement tests for literacy, language, and math as well as teacher ratings of classroom behavior. The pre-K participants and nonparticipants in this comparison were virtually identical on baseline variables such as achievement pretests, demographics, and family characteristics.

We found statistically significant positive effects for the pre-K participants at the end of the pre-K year on all of the achievement measures except one (oral comprehension), as well as on the composite achievement score. In addition, benefits were observed on teacher ratings obtained at kindergarten entry on three measures: preparedness for grade, work-related skills, and social behavior.

Regression discontinuity design studies. The short-term effects of state pre-K programs have most often been studied using a research design that takes advantage of age thresholds for enrollment in state pre-K programs. Children entering kindergarten who attended a pre-K program in the previous year (treatment group) are assessed in the fall of their kindergarten year, and their results are compared with the results of children who, because of their later birthdays, are just entering the pre-K program at that time (control group). With statistical adjustments for the age difference, estimates of the pre-K effects can be derived. This method is called an *age-cutoff regression discontinuity design* (RDD); because the treatment and control groups are not created via random assignment, it is considered a quasi-experimental design.

The RDD is widely acknowledged to be one of the more rigorous quasi-experimental designs.[20] What those using this design typically do not acknowledge is that the age-cutoff version is not a true RDD but, rather, an approximation that is vulnerable to biases that can compromise the effect estimates.[21] For example, children from economically disadvantaged families who attend pre-K are more likely to move[22] and thus be omitted from outcome assessments conducted at the beginning of kindergarten the next year. Also, the children in the control group who are assessed at the beginning of pre-K were 3 years old the previous year. In that regard, they are not comparable to the pre-K participants assessed at the beginning of kindergarten who were 4 years old the previous year. A year lived as a 3-year-old provides different background experiences than a year in the life of a comparable 4-year-old would have.

The majority of pre-K age-cutoff RDD studies have been conducted by researchers associated with the National Institute for Early Education Research and include state-funded pre-K programs in Arkansas,[23] California,[24] Michigan,[25] New Jersey,[26] New Mexico,[27] Oklahoma,[28] South Carolina,[29] and West Virginia.[20,30] The validity of the results reported for these studies is difficult to assess—they provide little detail about the equivalence of the children in the successive pre-K cohorts that are compared, the data used in the analyses, or the nature of the analyses. Other age-cutoff studies conducted in Georgia,[31] North Carolina,[32] and Tennessee,[33] as well as the earliest of such studies done in Oklahoma,[34] have used statistical controls to better match the comparison groups on such characteristics as gender, race, and English language proficiency.

The outcomes measured in the age-cutoff RDD studies have almost exclusively been cognitive measures, for example, for emerging literacy, language, and math knowledge. Overall, the results on such measures show wide variation but almost always indicate positive pre-K effects. Only the Georgia study included behavioral outcomes, in particular for social skills, problem behavior, and social awareness, and positive effects were found on those as well.

Post hoc matched studies. A number of studies use a design that compares the results of outcome measures taken at the beginning of the kindergarten year by children who participated in the state pre-K program the previous year with those for children who did not attend the program. These children are typically matched on a few demographic variables, such as gender, race or ethnicity, age, and FRPL eligibility. However, these post hoc matched studies lack critical information about how comparable the children were on such factors as initial cognitive skills, relevant behavioral dispositions, and family background prior to their differential exposure to pre-K.

Without "before" measures, we cannot know if differences in the outcomes were a result of pre-K participation or preexisting differences between the groups. An inherent difference between these groups is that the parents of pre-K participants chose to send their children to a voluntary educational program, whereas the parents of nonparticipants did not. The greater motivation or ability of the parents who enrolled their children in pre-K, compared with that of the parents who did

not, likely reflects family differences that could easily be related to children's academic performance.

Post hoc matched studies have been reported for Arkansas,[23] California,[24] Michigan,[25] New Jersey,[35,36] and Virginia.[37] Virtually all of the effects reported in these studies favor the pre-K participants, although they are generally smaller than those found in the RDD studies, even in studies of the same state programs conducted by the same researchers (e.g., in Arkansas and New Jersey). These studies are easy to do, but this posttest-only design is weak and has long been recognized as a preexperimental design that is incapable of supporting causal inferences about intervention effects.[38]

Albeit based largely on methodically weak study designs, rather consistent evidence of short-term benefits has been found for state pre-K programs, as described above. Moreover, that conclusion is consistent with the broader research literature on the short-term cognitive effects of early childhood education.[18] However, the available research on the extent to which those positive effects last past the end of the pre-K year is far shakier.

Long-Term Effects

Randomized studies. Our RCT study in Tennessee is, again, the only one of its kind that reports longer term outcomes of pre-K program participation.[21] We have followed the subsample of children from our study through third grade and found that, despite the positive achievement gains made by the pre-K participants relative to the gains of the nonparticipants during the pre-K year, there was no longer any difference in those outcomes by the end of kindergarten. Moreover, by the end of third grade, the control group of children who did not attend pre-K actually scored higher than the pre-K participants did on some achievement measures. That is, we observed an acceleration in achievement for pre-K nonparticipants such that their performance soon caught up with, and in some cases surpassed, that of the pre-K participants. By the end of first grade, teacher ratings of preparedness for grade, work-related skills, social behavior, and related noncognitive outcomes similarly showed no differences between the pre-K participants and nonparticipants.

Post hoc matched studies. The largest number of studies of longer term effects of state pre-K programs

use post hoc matched designs to compare outcomes for former pre-K participants and nonparticipants some years later. These designs have the same sources of potential bias described earlier, plus the added possibility of differential attrition from the sample as time goes on. Such comparisons have been reported for state pre-K programs in Arkansas,[23,39] Colorado,[40] Louisiana,[41] Michigan,[25,42] North Carolina,[43] New Jersey,[35,44] South Carolina,[45] Tennessee,[46] Texas,[47,48] Virginia,[37] and Washington.[49] The outcomes examined—mostly achievement test scores and grade retention—are almost exclusively drawn from state databases.

Overall, the results vary considerably in magnitude and statistical significance, or how likely the findings are due to chance, but they overwhelmingly favor the children who attended the state pre-K programs. Further, those differences are typically described in the language of causal attribution—for example, as *pre-K effects*—when that interpretation is not justified as a conclusion from such methodologically deficient research designs.

Consider, for example, the report prepared for the Louisiana Department of Education by a team of researchers from the University of Louisiana, the University of Alabama, and Georgetown University.[50] The state achievement test scores of third through eighth grade children eligible for the FRPL programs who had attended the state pre-K program were compared with the scores of FRPL-eligible students in the same grades who had not attended the state pre-K program. The higher scores for pre-K participants were interpreted as proof of pre-K having a "positive impact" on achievement. Yet no data were presented to establish that economically disadvantaged pre-K participants and nonparticipants were equivalent at the beginning of the pre-K year or, indeed, in any year on other factors that might have influenced their test performance.

Some of the post hoc matched studies obtained more data about student characteristics than did the Louisiana study, and those data were then used for matching or statistical control. However, only a few of these studies went beyond FRPL status, age, gender, and race or ethnicity, and none included any true baseline data beyond static demographics. Although these studies have been cited as evidence of sustained effects from state pre-K programs, they fall well below the most minimal methodological standards required to support even a tentative claim of that sort.

Difference-in-difference studies. Another method that researchers have used to study the effects of state pre-K programs on school achievement is known as a *difference-in-difference* (DD) *design*. These studies examine the differences in state- or county-level student scores on measures administered after a pre-K program is rolled out and compare them with differences in test scores seen over a comparable period for another area in which there was no analogous pre-K implementation or expansion. The challenge for researchers using this design is to develop statistical analyses that isolate the difference made in the target outcomes by pre-K implementation from all of the other influential factors co-occurring over that same time period that are not necessarily also occurring in the same way in the comparison area.

For example, a DD design was used to investigate the effects of the Georgia universal pre-K program that grew from participation rates of 14% in 1995 to 55% in 2008.[51] The analysis included statistical controls for a range of factors other than the introduction of pre-K that could have influenced student outcomes. Initial analyses indicated positive pre-K effects on achievement scores from the National Assessment of Educational Progress, but further analyses exploring control group variants and different statistical models did not yield a robust conclusion. Similar sensitivity to including different kinds of information in the analyses and the use of alternate statistical models was found in the results for another DD study of the Georgia program as well as for the Oklahoma program.[52]

More robust findings emerged from a DD study of two early childhood programs in North Carolina, of which one (More at Four) was a pre-K program.[53] This study focused on differences across counties in the timing and magnitude of funding for these programs and was distinctive in at least two ways. First, a high proportion of the pre-K programs were not school based. Second, by drawing on birth records, researchers found an unusually rich set of statistical control variables for the analysis. Positive effects were found on third grade state achievement test scores for both reading and math in counties implementing pre-K early. The strong statistical control notwithstanding, the authors acknowledged that the validity of the findings rested on the assumption that there were no uncontrolled factors capable of influencing test scores that

were coincident with the increases in funding for the pre-K program.

The difficulty of drawing firm conclusions from DD analyses in the dynamic context of state pre-K expansion is further illustrated by an ambitious study conducted by Rosinksy.[54] She compared the 2007, 2009, and 2011 fourth grade National Assessment of Educational Progress math scores with program enrollment of 4-year-olds 6 years previously in Head Start, state-funded pre-K, or special education preschools across multiple states. Surprisingly, she found a negative association between math scores and higher enrollment in state-funded pre-K.

Although DD studies in general are methodologically stronger than the post hoc matching studies, the inconsistency in their findings makes it difficult to draw confident conclusions from them about the long-term effects of state pre-K programs.

Concluding observations about long-term pre-K effects. Overall, the methodological quality of research on the effects of state pre-K programs is poor, showing little improvement since a 2001 review[55] by Gilliam and Zigler that concluded that the knowledge base was dangerously weak just as states were ramping up their programs. The evidence for favorable immediate effects of pre-K participation is consistent across a range of research designs, despite their respective limitations. By contrast, not only is there little consistency in the findings on long-term effects, but few studies address that issue with strong research designs.

The limited range of outcome variables represented across all of the studies of state pre-K programs is also noteworthy. The measures used are almost exclusively indices of academic achievement, primarily achievement tests. Very few studies have examined behavioral outcomes, despite the fact that those are the pre-K effects that are supposed to be sustained in the long term and generate the cost savings that have been claimed for pre-K. Investigation of adult behavioral outcomes may be beyond the reach of time-limited state pre-K studies, but that is not the case for potential bridging variables that may connect pre-K experiences with adult outcomes, for example, self-regulation, engagement with school, and grit. Outcomes of this sort are strikingly absent from the research on state pre-K effects.

Problems in Taking Programs to Scale

The research community has developed and validated many promising programs and practices, but few of these have been taken to scale while maintaining the same level of effectiveness.[56] In the case of pre-K programs implemented statewide, the situation is especially problematic. What is being scaled up is not a well-defined practice but rather an idea, a concept— the notion that some kind of school-like intervention provided to poor children prior to kindergarten entry will change their developmental trajectories in positive ways that will last well into adulthood. The grounding for this expectation comes from research conducted on the Cadillac programs implemented half a century ago that bear little resemblance to today's state programs.

In this regard, the recent scale-up effort for Response to Intervention (RTI; see the online Supplemental Material for more information) is instructive. RTI was developed and favorably evaluated in small experimental studies of a closely monitored small-group reading program for students in early grades. On that evidence, the program was written into the reauthorization of the Individuals with Disabilities Education Act in 2004. However, RTI is essentially the set of concepts that guided the original experimental intervention rather than a structured program, and school districts were free to implement it however they saw fit. A multistate evaluation of RTI implemented at scale actually found negative effects.[57]

The pre-K concept is much like the situation with RTI: There is no coherent vision for what the program should look like other than it should somehow be like the Perry Preschool Project and the Abecedarian Project—but not so much so that it would require as much time, effort, staff, or money. Viewed realistically, policymakers should not expect a scaled-up version of this rather open-ended concept to have the dramatic, sustained effects that are forecast by advocates and proponents.[58] As the research we have reviewed above indicates, little evidence supports the expectation of such effects.

Conclusion

The importance of helping young children from impoverished backgrounds advance in school learning and behavior cannot be overstated. The idea that a year

of pre-K can have effects that will endure well into adulthood is appealing to policymakers, school administrators, businessmen,[59] and law enforcement officials.[60] If the achievement gap for poor children can be closed and life outcomes improved with a year of preschool, then state policymakers would be foolish not to implement programs making pre-K accessible to everyone. This idea and the children's needs are so very compelling that states have rushed to scale up pre-K programs without much attention to the question of how to design and support those programs so that they are effective.

Much of the research reviewed above was conducted or commissioned by the state departments of education that administer the respective pre-K programs. That the resulting reports do not mention the methodological weaknesses of these studies while they emphasize the positive effects the studies are purported to demonstrate raises a question about the purpose of this research. Many reports have the appearance of supporting state policies that have already been adopted. If the report writers adopted a more critical approach to describing such studies, the reports policymakers base their decisions on would not only be more forthright about the methodological limitations of the studies and less rosy about their conclusions, but they would also acknowledge the considerable difficulty of implementing an effective program at scale and avoid claiming or implying that scale-up had been successfully accomplished.

It is facile to call for more and better research, but the problems outlined here are fundamental and must be addressed. Even if good, long-lasting effects were found for some pre-K programs, researchers would be hard-pressed to specify which features of those programs were responsible for that success so they could be emulated elsewhere. What is lacking is a distinct and plausible *theory of change* that would articulate the expectations for what should happen in pre-K and how that would affect the desired immediate and long-term outcomes. Such a theory would help identify the outcomes states should monitor and the adjustments that might be tried if those outcomes are not satisfactory. Moreover, the expectations for state pre-K emphasize long-term effects, but policymakers are not prepared to wait 15 to 20 years to find out if their programs are successful. Researchers need to identify and validate the proximal outcomes in a theory

of change that are the precursors to the desired long-term outcomes.

Children are not well served by a perpetuation of magical thinking about the likelihood of profound effects resulting from poorly defined state-run pre-K programs. Moreover, researchers should not be aiding and abetting that thinking with weak and misleading research presented without acknowledgement of its serious limitations. Viewed with a critical eye, the currently available research raises real questions about whether most state pre-K programs do anything more than boost 4-year-olds' academic cognitive skills to where they would be by the end of kindergarten anyway.

author affiliation

Farran, Department of Teaching and Learning, Vanderbilt University; Lipsey, Department of Human and Organizational Development, Vanderbilt University. Corresponding author's e-mail: Dale.Farran@Vanderbilt.edu

supplemental material

- https://behavioralpolicy.org/journal/
- Supplemental Text

References

1. Heckman, J. J. (2006, June 30). Skill formation and the economics of investing in disadvantaged children. *Science, 312,* 1900–1902. Retrieved from http://www.jstor.org/stable/3846426
2. Executive Office of the President of the United States. (2014, December). *The economics of early childhood investments.* Washington, DC: Author. Retrieved from https://www.whitehouse.gov/sites/default/files/docs/early_childhood_report1.pdf
3. Farley, R. (2013, February 20). *Obama's preschool stretch.* Retrieved from http://www.factcheck.org/2013/02/obamas-preschool-stretch/
4. Murray, C. (2013, February 20). *The shaky science behind Obama's universal pre-K.* Retrieved from http://www.bloombergview.com/articles/2013-02-21/the-shaky-science-behind-obama-s-universal-pre-k
5. Lee, M. Y. H. (2015, April 20). Obama's claim that every dollar spent on pre-kindergarten earns '$7 back.' *Washington Post.* Retrieved from http://www.bloombergview.com/articles/2013-02-21/the-shaky-science-behind-obama-s-universal-pre-k
6. Michigan Department of Education. (n.d.). *Great Start Readiness Program.* Retrieved from http://www.michigan.gov/documents/mde/GSRP-Overview_410757_7.pdf
7. Pianta, R., & Wolcott, C. (2014). *Pre-kindergarten for the modern age: A scalable, affordable, high-quality plan for*

Texas. Austin, TX: Raise Your Hand Texas. Retrieved from http://www.raiseyourhandtexas.org/wp-content/uploads/2015/02/RaiseYourHandTexas_PreK_ResearchReport_Aug2015.pdf

8. Reardon, S. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 91–116). New York, NY: Russell Sage Foundation.

9. Farran, D. C. (2007). Is education the way out of poverty? A reflection on the 40th anniversary of Head Start (with commentaries by James King and Bernard L. Charles). *Monographs of the Center for Research on Children's Development and Learning, 3*.

10. Klein, J. (2011, July 7). Time to ax public programs that don't yield results. *TIME Magazine.* Retrieved from http://content.time.com/time/nation/article/0,8599,2081778,00.html

11. Farran, D., & Lipsey, M. (2015, October 8). Expectations of sustained effects from scaled up pre-K: Challenges from the Tennessee study. *Evidence Speaks Reports, 1*(3). Washington, DC: Brookings Institution.

12. Quinton, S. (2015, September 4). States agree on need for 'preschool,' differ on definition [Blog post]. Retrieved from http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2015/09/04/states-agree-on-need-for-preschool-differ-on-definition

13. Connors-Tadros, L., Brown, K., & Schilder, D. (2015). *Examples of state guidance to determine and verify income eligibility for prekindergarten programs.* Rutgers, NJ: Center for Enhancing Early Learning Outcomes.

14. U.S. Department of Education & U.S. Department of Health and Human Services. (2014, August 18). *Applications for new awards; Preschool development grants–development grants. Federal Register, 79*(159), 48853–48872.

15. What Works Clearinghouse. (2015). *Find what works/early childhood interventions.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved November 7, 2015, from ies.ed.gov/ncee/wwc/findwhatworks.aspx#Social-emotionaldevelopment

16. Duncan, G., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27,* 109–131. doi:10.1257/jep.27.2.109

17. Heckman, J., Moon, H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics, 94,* 114–128. doi:10.1016/j.jpubeco.2009.11.001

18. Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112,* 579–620.

19. Lipsey, M. W., Farran, D. C., & Hofer, K. (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade.* Retrieved from Peabody Research Institute website: http://peabody.vanderbilt.edu/research/pri/VPKthrough3rd_final_withcover.pdf

20. Wong, V. C., Cook, T. D., Barnett, W. S., & Jung, K. (2008). An effectiveness-based evaluation of five state pre-kindergarten programs. *Journal of Policy Analysis and Management, 27,* 122–154. doi:10.1002/pam.2031

21. Lipsey, M., Weiland, C., Yoshikawa, H., Wilson, S., & Hofer, K. (2015). Prekindergarten age-cutoff regression-discontinuity design: Methodological issues and implications for application. *Educational Evaluation and Policy Analysis, 37,* 296–313. doi:10.3102/0162373714547266

22. Coulton, C., Theodos, B., & Turner, M. (2012). Residential mobility and neighborhood change: Real neighborhoods under the microscope. *Cityscape: A Journal of Policy Development and Research, 14,* 55–89.

23. Jung, K., Barnett, W. S., Hustedt, J. T., & Francis, J. (2013). *Longitudinal effects of the Arkansas Better Chance Program: Findings from first grade through fourth grade.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

24. Barnett, W. S., Howes, C., & Jung, K. (2009). *California's state preschool program: Quality and effects on children's cognitive abilities at kindergarten entry* [Report to the California Children and Families Commission]. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

25. Xiang, Z., & Schweinhart, L. J. (2002). *Effects five years later: The Michigan School Readiness Program evaluation through age 10* [Report to the Michigan State Board of Education]. Retrieved from HighScope Educational Research Foundation website: http://www.highscope.org/file/Research/Effects%205%20Years%20Later.pdf

26. Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of New Jersey's Abbott Preschool Program on young children's school readiness.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

27. Hustedt, J. T., Barnett, W. S., Jung, K., & Friedman, A. H. (2010). *The New Mexico pre-K evaluation: Impacts from the fourth year (2008–2009) of New Mexico's state-funded pre-K program.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

28. Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of Oklahoma's early childhood four-year-old program on young children's school readiness.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

29. Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of South Carolina's early childhood programs on young children's school readiness.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

30. Lamy, C., Barnett, W. S., & Jung, K. (2005). *The effects of West Virginia's early education program on young children's school readiness.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

31. Peisner-Feinberg, E. S., Schaaf, J. M., LaForett, D. R., Hildebrandt, L. M., & Sideris, J. (2014). *Effects of Georgia's pre-K program on children's school readiness skills: Findings from the 2012–2013 evaluation study.* Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute.

32. Peisner-Feinberg, E. S., & Schaaf, J. M. (2011). *Summary of key findings: Effects of the North Carolina More at Four prekindergarten program on children's school readiness skills.* Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute.

33. Lipsey, M. W., Farran, D. C., Bilbrey, C., Hofer, K. G., & Dong, N. (2011). *Initial results of the evaluation of the Tennessee Voluntary Pre-K program.* Nashville, TN: Vanderbilt University, Peabody Research Institute. Retrieved from https://my.vanderbilt.edu/tnprekevaluation/files/2013/10/April2011_PRI_Initial_TN-VPK_ProjectResults.pdf

34. Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41,* 872–884.

35. Frede, E., Jung, K., Barnett, W. S., & Figueras, A. (2009). *The APPLES blossom: Abbott Preschool Program longitudinal effects study (APPLES). Preliminary results through 2nd grade.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

36. Frede, E., Jung, K., Barnett, W. S., Lamy, C. E., & Figueras, A. (2007). *The Abbott Preschool Program longitudinal effects study (APPLES): Interim report.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

37. Huang, F. L., Invernizzi, M. A., & Drake, E. A. (2012). The differential effects of preschool: Evidence from Virginia. *Early Childhood Research Quarterly, 27,* 33–45.

38. Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin.

39. Hustedt, J. T., Barnett, W. S., & Jung, K. (2008). *Longitudinal effects of the Arkansas Better Chance Program: Findings from kindergarten and first grade*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

40. Colorado Department of Education. (2012). *Colorado preschool program: 2012 legislative report*. Denver, CO: Author.

41. Picard Center for Child Development and Lifelong Learning. (2013). *Eighth-grade outcomes for LA 4 cohort 1 students* [Technical brief]. Lafayette, LA: Author.

42. Malofeeva, E. V., Daniel-Echols, M., & Xiang, Z. (2007). *Findings from the Michigan School Readiness Program 6 to 8 follow up study*. Retrieved from HighScope Educational Research Foundation website: http://www.highscope.org/file/Research/6%20-%208%20follow%20up%20report%20FINAL%2010-9-07.pdf

43. Peisner-Feinberg, E. S., & Schaaf, J. M. (2010). *Long-term effects of the North Carolina More at Four prekindergarten program: Children's reading and math skills at third grade*. Chapel Hill: University of North Carolina, Frank Porter Graham Child Development Institute.

44. Barnett, W. S., Jung, K., Youn, M.-J., & Frede, M. C. (2013). *Abbott Preschool Program longitudinal effects study: Fifth grade follow-up*. New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.

45. Barnett, W. S., Frede, E. C., Mobasher, H., & Mohr, P. (1987). The efficacy of public preschool programs and the relationship of program quality to efficacy. *Educational Evaluation and Policy Analysis, 10*(1), 37–49.

46. Strategic Research Group. (2011). *Assessing the impact of Tennessee's pre-kindergarten program: Final report*. Columbus, OH: Author.

47. Andrews, R. J., Jargowsky, P., & Kuhne, K. (2012). *The effects of Texas's targeted pre-kindergarten program on academic performance* (CALDER Working Paper No. 84). Washington, DC: American Institutes for Research.

48. Huston, A., Gupta, A., & Schexnayder, D. (2012). *Study of early education in Texas: The relationship of pre-K attendance to 3rd grade test results*. Austin, TX: University of Texas, Ray Marshall Center for the Study of Human Resources.

49. Bania, N., Kay, N., Aos, S., & Pennucci, A. (2014). *Outcome evaluation of Washington State's Early Childhood Education and Assistance Program* (Document No. 14-12-2201). Olympia: Washington State Institute for Public Policy.

50. Picard Center for Child Development and Lifelong Learning. (2007–2008). *LA 4 & starting points: Prekindergarten program evaluation*. Lafayette, LA: Author

51. Fitzpatrick, M. (2008). Starting school at four: The effect of universal pre-kindergarten on children's academic achievement. *The B.E. Journal of Economic Analysis & Policy, 8,* 1–38.

52. Cascio, E. U., & Schanzenbach, D. W. (2013, September). *The impacts of expanding access to high-quality preschool education*. Paper presented at the Fall 2013 Conference on the Brookings Papers on Economic Activity, Washington, DC. Retrieved from http://www.brookings.edu/about/projects/bpea/papers/2013/fall-cascio-preschool-education

53. Ladd, H. F., Muschkin, C. G., & Dodge, K. A. (2014). From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina. *Journal of Policy Analysis and Management, 33,* 162–187.

54. Rosinsky, K. (2014). *The relationship between publicly funded preschool and fourth grade math test scores: A state-level analysis* (Master's thesis, Georgetown University). Retrieved from https://m.repository.library.georgetown.edu/bitstream/handle/10822/709852/Rosinsky_georgetown_0076M_12517.pdf?sequence=1&isAllowed=y

55. Gilliam, W. S., & Zigler, E. F. (2001). A critical meta-analysis of all evaluations of state-funded preschool from 1977 to 1998: Implications for policy, service delivery and program evaluation. *Early Childhood Research Quarterly, 15,* 441–473.

56. Granger, R. C. (2011, Winter). The big why: A learning agenda for the scale-up movement. *Pathways,* 28–32.

57. Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015). *Evaluation of Response to Intervention practices for elementary school reading* (NCEE 2016-4000). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

58. Kirp, D. (2015, October 3). Does pre-K make any difference? *The New York Times.* Retrieved from http://www.nytimes.com/2015/10/04/opinion/sunday/does-Pre-K-make-any-difference.html?_r=0

59. ReadyNation. (n.d.). *Business case for early childhood investments*. Retrieved from http://readynation.s3.amazonaws.com/wp-content/uploads/ReadyNation-Business-Brief-Endnotes.pdf

60. Christeson, W., Bishop-Joseph, S., O'Dell-Archer, N., Beakey, C., & Clifford, K. (n.d.). *I'm the guy you pay later.* Retrieved from Fight Crime: Invest in Kids website: http://cdn.fightcrime.org/wp-content/uploads/I'm_The_Guy_Report.pdf

Reforming Head Start for the 21st century: A policy prescription

Sara Mead & Ashley LiBetti Mitchel

Summary. Head Start was born in 1965 as a federal program that aimed to lift America's neediest children out of poverty and enhance their lifetime opportunities. Today, Head Start continues to play an important role in our nation's early learning and development system; it serves nearly 1 million children and remains the only preschool option for poor children in many communities. Yet Head Start faces real challenges if it is to remain relevant and competitive in the face of the surge in state-funded prekindergarten (pre-K) programs over the past 25 years. State pre-K programs now serve 1.3 million children and typically spend about half the amount per child that Head Start does, yet the best state pre-K programs achieve better results than does the average Head Start program. And recent federally funded evaluations of Head Start raise serious questions about its long-term effectiveness. In this article, we examine the major actions undertaken by bipartisan policymakers to improve Head Start and propose three distinct prescriptions of our own: (a) Allow Head Start providers and grantees the flexibility to triage the services most needed by children in their program rather than follow the "all services to all kids" mandate that now exists, (b) shift performance measures to focus more on outcomes than on compliance with regulations, and (c) change federal policies so that Head Start grantees can more easily coordinate and integrate with local and state early education services and funding streams.

Policymakers have been debating Head Start's effectiveness for nearly as long as the program has existed. In 1969, a study by the Westinghouse Learning Corporation, a research organization, found that Head Start produced gains in cognitive and language skills at the end of first grade but that these gains "faded out" by the end of second and third grades. Despite flaws in the study's methodology, the study propagated the idea that Head Start does not work. And the government's own 2010 study of Head Start—the Head Start Impact Study (HSIS), a rigorous, federally funded evaluation—found that although Head Start students made meaningful

gains in early literacy and the program enhanced some behavior, health, and parenting skills in participating Head Start families, as compared with a control group of nonparticipants, the gains quickly evaporated; in this case, the advantage was gone by the time kids completed first grade.

In 2014, future House Speaker Paul Ryan's antipoverty budget plan stated, "Although certain Head Start centers have made a positive difference for select beneficiaries, the program overall has a disappointing record."[1] Checker E. Finn, Jr., a former U.S. assistant secretary of education who currently conducts policy analysis for prominent think tanks, has frequently criticized the program, stating, "Despite its popularity, despite the billions spent on it, and notwithstanding the decent job it does of targeting services on needy kids, today's Head Start, when viewed through the lens of pre-K education and kindergarten readiness, amounts to a wasted opportunity."[2] And *TIME Magazine*, citing the disappointing results to the HSIS, called for an end to the program.[3] Others are more optimistic: "Weighing all of the evidence . . . , the most accurate conclusion is that Head Start produces modest benefits including some long-term gains for children," said W. Steven Barnett, an education professor at Rutgers University in an op-ed for the *Washington Post* in 2013.[4]

Disheartening as the initial results of the HSIS were to policymakers, there is evidence from further analysis of the HSIS data that kids enrolled in Head Start had stronger vocabulary and cognitive outcomes than did kids who stayed at home with a parent or relative during these early years and that their parents reported fewer child behavior problems; these gains were sustained through elementary school. There was little difference between results for Head Start students and kids in other center-based child care or preschool programs.[5,6] In other words, although Head Start may not perform better than other preschool programs, it is better than not attending preschool at all. And many of the children Head Start serves would not otherwise attend preschool.

Even given the significant expansion of state prekindergarten (pre-K) programs this century and the end of the last, there are far fewer state pre-K slots than there are low-income children,[7] and eight states do not offer publicly funded preschool at all.[8] As a result, only 60% of poor 4-year-olds attend preschool.[9]

But other evidence (see Table 1) suggests that Head Start could do better than it currently does. Studies of high-quality, publicly funded state pre-K programs that serve significant numbers of children in New Jersey, Boston, and Tulsa find evidence of learning gains at kindergarten entry that are larger than those found in the Head Start Impact Study, with some of these gains lasting well into the elementary years.[10] Moreover, Head Start effects vary significantly across centers—even more than they vary across measures of kindergarten through 12th grade (K–12) school effectiveness.[11] In other words, some Head Start programs produce dramatically better results than others.

> *Some Head Start programs produce dramatically better results than others.*

Given this evidence, the relevant question for policymakers is not whether Head Start works but how to increase the number of Head Start centers that work as well as the most effective Head Start centers and state-funded pre-K programs.

Bipartisan Efforts to Improve Head Start

Federal policymakers have taken numerous steps to improve the quality and impact of Head Start through the required reauthorizations of the program. In 1998,

Table 1. Comparison of two public high-quality prekindergarten programs to Head Start

Measure	Boston Preschool Program	New Jersey Abbott Preschool program	Head Start Impact Study
Impacts found at end of preschool year[a]	Positive impacts on vocabulary, early reading, numeracy, and social-emotional skills	Positive impacts on vocabulary, math, and print awareness skills	Positive impacts on prereading, prewriting, and vocabulary skills and parent reports of children's literacy skills
Effect size[b]	.44−.62	.40	.10−.34
Gains lasted through	3rd grade	5th grade	No gains found after first grade for full sample
Cost per child	$12,000	$12,000−$14,900	$10,526[c]

Note. The data from this table come from the following sources: "Early Learning: The New Fact Base and Cost Sustainability," by J. Minervino and R. Pianta, 2013, https://docs.gatesfoundation.org/documents/Lessons%20from%20Research%20and%20the%20Classroom_September%202014.pdf; *Head Start Impact Study: First Year Findings,* by the U.S. Department of Health and Human Services, Administration for Children and Families, June 2005, http://www.acf.hhs.gov/sites/default/files/opre/first_yr_finds.pdf; and *The Effects of New Jersey's Abbott Preschool Program on Young Children's School Readiness*, by C. Lamy, W. S. Barnett, and K. Jung, December 2005, http://nieer.org/resources/research/multistate/nj.pdf.

[a]The comparisons are between program participants and nonparticipants—that is, kids with no preschool experience or those who attended another program. [b]Effect sizes provide a quantitative way of comparing the relative strength of effects found in different studies. In the social sciences, an effect size of .2 is generally considered small, an effect size of .5 is considered medium, and an effect size over .8 is considered large. Thus, the effect sizes found for the Boston and Abbott programs are generally medium, whereas those found for key indicators in the Head Start Impact Study were small. [c]There is no set per-child cost in Head Start, but on average, the federal government spent $8,771 per Head Start child served in fiscal year 2015. Including the required 20% nonfederal match, Head Start programs can be assumed to spend an average of $10,526 per child in federal and nonfederal funds, although actual spending levels vary widely.

the Coats Human Services Amendments required Head Start to develop education performance standards and increased educational requirements for Head Start teachers.[12] The Improving Head Start for School Readiness Act of 2007 expanded on these reforms, requiring all Head Start teachers to have an associate's degree by 2011 and half to have a bachelor's degree by 2013.[13] The 2007 act also mandated that monitoring of Head Start grantees include measures of teaching quality, using "valid and reliable" observations of adult−child interactions.[14] The Department of Health and Human Services (HHS) Office of Head Start subsequently adopted the Classroom Assessment Scoring System (CLASS), an observational tool developed by researchers to measure these interactions. The 2007 act also sought to increase coordination between Head Start, state preschool programs, and local public schools by, among other things, mandating that Head Start work with local school districts to serve children with disabilities and prepare children to transition to kindergarten. The act also funded the creation of State Advisory Councils on Early Childhood Education and Care, which bring together representatives from various state agencies that serve young children and their families, as well as other early childhood providers and stakeholders, to advise on early childhood policies and improve coordination across programs.[15]

The Designation Renewal System

Also, under the 2007 act, Head Start grants changed from continuous grants, with no expiration date, to renewable, 5-year grants. Grants are automatically renewed if federal monitoring finds that Head Start grantees are operating a "high-quality and comprehensive Head Start program." But grantees that are found to have one of seven red-flag criteria (listed below) are required to compete to renew their grant—a process known as *designation renewal*. During the designation renewal process, other organizations can also apply for the grants, which could result in the incumbent losing the grant. The Office of Head Start uses the following seven criteria to identify grantees that must compete:

- deficiencies identified through Head Start's monitoring system, through which monitors conduct site visits and review documentation to ensure programs are complying with Head Start performance standards;

- license revocations by state or local child-care licensing agencies;
- suspensions by the Department of HHS;
- disbarment by any other state or federal agencies;
- significant financial risk;
- failure to establish or track and analyze data on goals for improving children's school readiness; and
- CLASS scores that fall below a certain threshold or within the lowest 10% of grantees on any of the three CLASS domains.[16]

As of July 2015, all current Head Start grantees have been evaluated under the designation renewal system. (For more information on the schedule of designation renewal and the criteria used for each cohort, see the online Supplemental Material.)

Roughly one-quarter of Head Start grantees were required to compete through designation renewal from 2012 to 2015. Of those, 22% lost their grants. Nearly 5% of Head Start grants changed hands over the past 3 years as a result of designation renewal.[17–20] (By way of comparison, the closure rate for charter schools—which were explicitly created on the assumption that schools that fail to improve performance should be closed—was about 3% per year over the same time period.)[21] This suggests that designation renewal has been effective in expelling underperforming grantees and allowing others to replace them, particularly when compared with rates of provider turnover in some other education and social service programs. Some large Head Start grants have also been split between a previous grantee and other agencies that had served children as its subcontractors. Breaking up large grants may help improve quality by giving former subcontractors more autonomy and control over resources and shifting resources closer to the level at which the child is actually served.

Limitations of the Designation Renewal System

Anecdotal reports suggest that designation renewal has motivated remaining grantees to address long-standing problems and focus on improving the quality of teaching. "Designation renewal had more influence on grantees changing what they do than the millions that the federal government invests in training and technical assistance," says one former federal official involved in implementing the designation renewal system, adding, "The thought of losing money moves people to change."

But the process also has limitations. The criteria for identifying grantees to compete may not be the right ones. They place a heavy weight on compliance with Head Start or other state and federal regulations[22] and too little on actual outcomes for kids and families. In recent designation renewal cycles, 44% of the grantees that participated in designation renewal were required to do so on the basis of their CLASS scores, but two-thirds of those also had other compliance issues. And 30% of those required to participate in designation renewal were identified on the basis of immediate self-reports—most commonly incidents in which a child was left unattended (which Head Start staff are required to report to regional offices). No grantee was required to compete on the basis of child or family outcomes. That is because the criteria for designation renewal do not include measures of child and family outcomes, and it illustrates the extent to which current policies incentivize providers to focus on compliance rather than improving outcomes for children and families. The result has been to intensify a culture within Head Start that focuses on adherence to bureaucratic rules rather than on how programs can better serve children and families, thereby reducing opportunities for innovation.

CLASS is a measure of program quality, but because it only measures one dimension of quality—teacher–child interactions—and is administered in only a sample of classrooms, it does not give a comprehensive picture of program quality. Current designation renewal criteria require grantees to compete if they score in the lowest 10% of all grantees on any of three CLASS domains—but on two of the three domains, the cutoff for the lowest 10% still reflects a relatively high level of quality. As a result, some programs required to compete because of their CLASS scores are likely delivering high-quality early learning experiences for children, while the criteria miss other programs that are not delivering quality learning experiences.

Designation renewal was meant to enable new providers to compete to replace existing providers, yet few applied. Of more than 250 organizations that received grants in the first two rounds of renewal, all but 13 already operated Head Start programs. The current grantee is often the only applicant, causing incumbent grantees inconvenience but providing no true competition. And sometimes the Office of Head Start receives no applications that meet the bar to receive funding. The lack of qualified applicants undermines the designation

Explaining CLASS

The Classroom Assessment Scoring System (CLASS) was selected by the Office of Head Start to provide a valid and reliable observational measure of quality in Head Start monitoring reviews, as required by the Improving Head Start for School Readiness Act of 2007.

CLASS, which evolved out of tools developed by researchers in the 1990s to evaluate child-care quality, focuses on the teacher–child interactions in early childhood settings, which research shows predict children's learning much better than traditional pre-K quality measures such as class size, adult-to-child ratios, and teacher qualifications.[A–C] Interactions in pre-K classrooms are rated in three domains:

- **Emotional Support** looks at teachers' sensitivity and responsiveness to children's perspectives and is related to children's social-emotional development.
- **Instructional Support** looks at the quality of instruction in classrooms, how teachers use language, and how they create experiences that foster children's learning. Scores on this domain of CLASS are correlated with children's cognitive development.
- **Classroom Organization** looks at how teachers structure the classroom, use time, and manage children's behavior.

All three domains of CLASS are scored on a 7-point scale. Scores above a 5 are considered good. Most early childhood programs perform relatively well on measures of emotional support and classroom organization and management but dismally on measures of instructional support.[D] This is true in Head Start as well.[E] In 2015, the average CLASS score for all Head Start grantees observed was 6.03 for emotional support, 2.88 for instructional support, and 5.8 for classroom organization.[F]

All CLASS reviewers complete extensive training and must pass a reliability test before being permitted to score programs for Head Start monitoring reviews.

[A]Pianta, R. C. (2007, Winter). Preschool is school, sometimes: Making early childhood education matter. *Education Next, 7*(1), 44–49. Retrieved from http://educationnext.org/preschool-is-school-sometime

[B]Pianta, R. C., Barnett, W. S., Burchinal, M., & Thornburg, K. R. (2009). The effects of preschool education: What we know, how public policy is or is not aligned with the evidence base, and what we need to know. *Psychological Science in the Public Interest, 10*(2), 49–88.

[C]Sabol, T. J., Hong, S. L. S., Pianta, R. C., & Burchinal, M. R. (2013, August 23). Can rating pre-K programs predict children's learning? *Science, 341,* 845–846. doi:10.1126/science.1233517s

[D]Early, D., Barbarin, O., Bryant, D., Burchinal, M., Chang, F., Clifford, R., . . . Barnett, S. (2005). *Pre-Kindergarten in eleven states: NCEDL's multi-state study of pre-kindergarten & Study of State-Wide Early Education Programs (SWEEP). Preliminary descriptive report* [Working paper]. Available from Frank Porter Graham Child Development Institute website: http://www.fpg.unc.edu/node/4654

[E]Office of Head Start, Early Childhood Learning and Knowledge Center. (n.d.). *Use of Classroom Assessment Scoring System (CLASS®) in Head Start.* Retrieved from https://eclkc.ohs.acf.hhs.gov/hslc/hs/sr/class/use-of-class.pdf

[F]Office of Head Start, Early Childhood Learning and Knowledge Center. (n.d.). *A national overview of grantee CLASS® scores in 2015.* Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/class-reports/docs/national-class-2015-data.pdf

renewal system as an accountability measure and threatens access to early learning in communities where Head Start is the only preschool option.

Why aren't more organizations applying for Head Start funds? Funding levels provided by Head Start may be too low to cover the cost of delivering quality programs or attract new applicants. The complexity of Head Start's requirements and the relatively short application timeline also likely dissuade many prospective applicants.

Designation renewal is based on sound ideas: accountability and an opportunity to replace weak providers with new blood. But the implementation of the process has had weaknesses, as outlined above. Improving the designation renewal system will require including additional measures of program quality and outcomes besides compliance and CLASS; making grantees compete when they fall below an absolute threshold for bad CLASS scores rather than when they are part of the lowest 10% of performers; and intentionally cultivating the supply of new applicants for Head Start grants where the incumbent provider is historically weak.

Revising the Head Start Performance Standards Under the Obama Administration

Head Start Performance Standards, the federal rules that govern the operation of Head Start programs,[23] address

everything from education, to parent engagement to finances. In September 2016, the Office of Head Start finalized a new version of the performance standards. This is the first major revision of the performance standards since 1998 and the first complete overhaul since their creation 40 years ago. Because the old standards have been amended piecemeal over the years, they were often confusing and redundant. The new standards are streamlined in number and clarity to make them easier to navigate, and they codify expectations for grantees. They also reflect recent research on how young children learn and on the characteristics of quality early childhood programs that have produced lasting learning gains in other studies. Key changes include the following:

- **Increasing dosage, or time spent in class:** The new standards more than double the minimum required hours for Head Start programs from 448 to 1020.[24] Currently, Head Start grantees are permitted to offer several preschool options—including full- or part-day programs, 4 or 5 days a week, and home-based preschool options—but only 43% of Head Start preschool programs offer classes that last 6 hours day, 5 days a week. Most of the rest offer only half-day programs, serve children 4 days a week, or both. The new standards phase in increased hour requirements: Head Start grantees must provide at least 1020 hours to 50% of children they serve by 2019 and 100% by 2021. They also allow programs to request the flexibility to offer a part-day program if doing so meets the community's needs. The administration proposed this change because research shows that kids who spend more time in early learning programs make greater learning gains; the relationship between time spent and learning is roughly proportional, meaning that children who spend twice as much time in preschool learn twice as much.[25,26] In the event Congress fails to appropriate the funds needed to implement the increased hour requirements, the Secretary of HHS may delay these added hour requirements.
- **Education and development:** The revised standards elevate the importance of the educational component of Head Start programs by providing more details about what quality early education

programs look like and emphasizing research-based practices in four core areas:
  - Teaching and the learning environment[27]
  - Curriculum
  - Child screening and assessment
  - Parent involvement
- **Parent engagement and support:** The new standards emphasize helping parents to support their children's learning by requiring programs to use a research-based parenting curriculum that focuses on building parents' confidence and skills to support their children's development and advocate for their children within the education system.[28] Although programs may continue to support families in other ways, such as by connecting them with social service and continuing education programs or helping them find housing and jobs, these changes clarify that the primary priority for Head Start family engagement is enabling parents to support children's learning and development.
- **Professional development:** The new standards shift the focus of professional development away from the one-shot workshops that are currently the norm to ongoing coaching. Research shows that this approach, in which a dedicated coach or center director regularly observes teachers in their classrooms (whether in person or by video), provides ongoing feedback, and helps teachers reflect on practice and set goals for improvement, results in better early childhood teaching and outcomes.[29,30]
- **Use of data:** The new standards require Head Start programs to collect, analyze, and use data to inform ongoing, continuous improvement. Effective early childhood programs regularly collect data—such as child learning outcomes; descriptive information on child and family demographics, well-being, and experiences; measures of teacher quality and professional development; and program operational data (such as financial indicators and staff turnover)—analyze that data to understand overall performance, trends, and opportunities for improvement; and make changes in practice in response to that analysis.[31] Practices for collecting and using data should be embedded into the practices of all program staff, from classroom teachers, to family support workers, to

center directors, to central leadership and board members. Reflecting this, the draft standards embed data and continuous improvement in program governance, education, child development, health, safety, and enrollment.

In revising the standards, the Administration for Children and Families eliminated a number of duplicative or overly prescriptive requirements and reduced the amount of paperwork for grantees. But current Head Start statute limits how much the standards can be streamlined, because it prohibits the secretary of the U.S. Department of Health and Human Services from reducing the range or scope of educational, health, dental, and other services that Head Start programs must provide.[32] And the new standards also impose new requirements on grantees on top of the above-noted revisions to current standards.

Implementing these new requirements will cost money, a lot of it. On average, the federal government spends about $8,800 per Head Start child served[33]—which is twice as much as the typical state pre-K program spends[34] but less than the costs of high-performing public preschool programs in Boston and New Jersey.[35] Federal regulatory impact analysis estimates a net cost of $1.05 billion to implement the new standards at current enrollment levels, with the longer day and year making up the lion's share of the price. The fiscal year 2016 omnibus appropriations increased Head Start funding by $570 million, including $294 million to extend the Head Start day.[36] But this increase will not cover the full cost to implement the standards.

Our Key Proposals

The adoption of CLASS, the designation renewal system, and the proposed revisions to the Head Start Performance Standards all represent improvements that support Head Start quality and outcomes. Yet further changes are needed to maximize the program's impact. On the basis of our past research and policy analysis on Head Start, our consulting work with high-performing Head Start grantees and other early childhood programs, and our review of research on effective early childhood practices and successful state pre-K programs, we propose the following changes.

Use a Triage System

Grantees should be allowed to determine the mix of family, health, and other services that are most important for preparing the children in their communities for school.

Current program standards require Head Start grantees to provide a wide variety of services, including early childhood education, family support services, nutrition services, oral health services, mental health services and referrals, and health screenings and referrals. This emphasis on comprehensive services, which has been a core feature of Head Start since its inception, reflects the integrated nature of child development and the complex, interrelated challenges facing poor families.

But over the past 50 years, a variety of federal and state programs—including Medicaid; the Affordable Care Act; community health centers; and the Women, Infants, and Children(WIC) food program—have been created to address many of these needs. Poor families may yet struggle to access services through these programs. But it is well worth asking whether Head Start programs should still be required to provide all currently mandated services for all children. Documenting currently required screenings, referrals, and other comprehensive services consumes considerable staff time and resources. Moreover, there is little evidence that early childhood programs offering these comprehensive services produce better results—either educationally or on health and other more holistic outcomes—than do those that focus on early learning. For example, an article published in 2010 analyzed the results of previous studies and found evidence of a negative relationship between programs' provision of comprehensive services and child outcomes.[37] Similarly, the HSIS found little evidence of improvements in most health outcomes for Head Start children, with the exception of dental health.[38]

That does not mean that Head Start should eliminate comprehensive services altogether. Many of the children Head Start services do need health, nutrition, or family supports to achieve their learning potential. But comprehensive service delivery in Head Start should be viewed as one tool in helping Head Start programs meet their ultimate goal: enabling children to enter kindergarten with the preparation and family support needed to succeed.

> *But comprehensive service delivery in Head Start should be viewed as one tool in helping Head Start programs meet their ultimate goal: enabling children to enter kindergarten with the preparation and family support needed to succeed.*

That means giving grantees greater flexibility to customize the family, health, and other services to reflect the needs of individual children and of the communities they serve rather than delivering all services to all children, as current rules require.

Develop Better Tools to Measure Program Performance and Child and Family Outcomes

Shift performance measures to focus more on outcomes rather than on compliance with regulations.
As we have shown, federal monitoring of Head Start focuses heavily on compliance with regulations. We believe the focus should shift to how well Head Start programs are preparing children to succeed in school and preparing their families to support this success. However, this shift cannot happen without better tools to assess children's learning and other developmental outcomes that support school readiness, as well as measures of program quality and impact on families.

Child development experts are understandably hesitant to place high stakes on measures of children's learning outcomes: Few existing measures of children's learning are valid, reliable, and sufficiently trusted by the field to be used in this way. Those measures that do exist often measure only a few components of development—and often not the most important outcomes.

To address this issue, Head Start officials should work with other federal agencies, researchers, grantees, and philanthropic groups to identify, develop, test, and refine new measures or improve existing ones, with the ultimate goal of adopting a set of measures that provides a comprehensive picture of Head Start performance at both the grantee and the program-wide levels. In the meantime, they can make better use of existing measures, such as setting cutoff points for unacceptable CLASS scores or establishing red flag indicators of child attendance and absenteeism. In the near term,

federal officials can also increase focus on performance by encouraging groups of grantees to adopt common measures of child learning and program quality, share data on these measures to compare their performance and identify strengths and weaknesses, and set individualized and collective goals for improvement. This approach creates incentives and support for programs to improve performance in areas that matter most, without mandating a single federal measurement tool. Groups of grantees could also pilot the range of new assessment tools that researchers are now developing—such as the early language screener developed by researchers at the Temple University Infant & Child Laboratory, the Lens on Science computer-adaptive science assessment, and the Minnesota Executive Function Scale—both to measure their performance and to determine whether these tools are appropriate for future program-wide adoption.

Facilitate Coordination with State Programs

Federal policies must provide flexibility and incentives for Head Start grantees to work with state and local programs.
Over the past 25 years, states and local school districts have dramatically expanded their involvement in early childhood education, creating state- and locally funded pre-K programs. They have also developed systems to measure the quality of early childhood programs, such as Quality Rating and Improvement Systems (QRIS), and created data systems to track information on children's participation in early childhood programs and link it with K–12 school data. (For more information on QRIS and data systems that link early childhood programs with K–12 data, see the online Supplemental Material.)

Head Start plays a central role in some of these efforts but has been excluded from others. Ensuring Head Start's future requires improving integration and coordination between Head Start and state and local early childhood initiatives. In our opinion, the federal government should not simply transfer funding and management of Head Start to the states, as some policymakers have proposed.[1] Rather, federal policymakers should focus on incentivizing Head Start, states, and local schools to work together and eliminate policies that prevent them from doing so. Working together

allows these entities to be more efficient, stream-lined, and integrated, reducing overall costs and the burden that families face when dealing with these systems separately.

For example, federal and state policies could:

- require Head Start to participate in state data systems (and require these systems to include Head Start);
- align federal and state standards for quality and child learning outcomes;
- allow state oversight (through QRIS or state pre-K programs) to take the place of some monitoring requirements for some Head Start grantees; or
- increase flexibility for grantees to braid together multiple state and federal funding streams to serve more children, lengthen the school day, or offer services in mixed-income settings.

In states with universal preschool access, Head Start funds might one day serve to supplement and enhance programming for at-risk children—as Title I funds do in K–12 schools—or shift to focus on infants and toddlers.

But any policies increasing the integration of Head Start and state pre-K must maintain or raise—not lower—quality standards for providers using Head Start funds and ensure that resources remain focused on the most at-risk children. This is complicated work that will likely require changes in state and local policies and programs, as well as in federal Head Start policies and grantee practices.

Conclusion

Over its 50-year history, Head Start has improved the lives of millions of children and their families by providing quality early learning programs and empowering parents to support their children's learning, and it continues to improve school readiness for our nation's most at-risk children. But Head Start needs additional changes. Policymakers must be willing to raise expectations for Head Start quality and outcomes, set clear priorities, find fair and accurate ways to measure programs' quality and effect on children's learning, and explore ways to better integrate Head Start with state and local preschool programs. These changes will be difficult, but they are necessary to maximize Head Start's impact for children and families.

author affiliation

Mead, partner at Bellwether Education Partners, a non-profit dedicated to helping education organizations become more effective in their work and achieve dramatic results, especially for high-need students. Mitchel, sr. policy analyst at Bellwether Education Partners. Corresponding author's e-mail: sara@bellwether education.org

References

1. House Budget Committee. (2014). *Expanding opportunity in America: A discussion draft from the House Budget Committee.* Retrieved from http://budget.house.gov/uploadedfiles/embargoed_expanding_opportunity_in_america___7232014.pdf
2. Finn, C. E., Jr. (2009, Fall). The preschool picture. *Education Next, 9*(4), 13–19. Available from http://educationnext.org/the-preschool-picture/
3. Klein, J. (2011, July 7). Time to ax public programs that don't yield results. *TIME Magazine.* Retrieved from http://content.time.com/time/nation/article/0,8599,2081778,00.html
4. Barnett, W. S. (2013, March 5). Does Head Start work for kids? The bottom line. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2013/03/05/does-head-start-work-for-kids-the-bottom-line/
5. Zhai, F., Brooks-Gunn, J., & Waldfogel, J. (2014). Head Start's impact is contingent on alternative type of care in comparison group. *Developmental Psychology, 50,* 2572–2586.
6. Feller, A., Grindal, T., Miratrix, L., & Page, L. (2014). *Compared to what? Variation in the impacts of early childhood education by alternative care-type settings* [Working paper]. Retrieved from http://scholar.harvard.edu/files/feller/files/feller_grindal_miratrix_page_12_6_14.pdf
7. Barnett, W. S. (2011, August). *Reforming early education* [Presentation notes]. Retrieved from Brookings Institution website: http://www.brookings.edu/~/media/events/2011/8/22%20early%20education/20110822_barnett_presentation.pdf
8. Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., & Squires, J. H. (2016). *The state of preschool 2015: State preschool yearbook.* New Brunswick, NJ: Rutgers University, National Institute for Early Education Research.
9. Barnett, W. S., & Yarosz, D. J. (2007). *Who goes to preschool and why does it matter?* (NIEER Preschool Policy Brief Issue 15). Retrieved from National Institute for Early Education Research website: http://nieer.org/resources/policybriefs/15.pdf
10. Bloom, H. S., & Weiland, C. (2015). *Quantifying variation in Head Start effects on young children's cognitive and socio-emotional skills using data from the National Head Start Impact Study.* Retrieved from http://www.mdrc.org/sites/default/files/quantifying_variation_in_head_start.pdf

11. Waters, C. (2014). *Inputs in the production of early childhood human capital: Evidence from Head Start* (NBER Working Paper No. 20639). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w20639.pdf

12. Coats Human Services Reauthorization Act of 1998, Pub. L. No. 105-285, 112 Stat. 2702 (1998).

13. Improving Head Start for School Readiness Act of 2007, 42 USC 9801 *et seq.*, § 648A (2007).

14. Improving Head Start for School Readiness Act of 2007, 42 USC 9801 *et seq.*, § 641A (2007).

15. Improving Head Start for School Readiness Act of 2007, 42 USC 9801 *et seq.*, § 642B (2007).

16. Improving Head Start for School Readiness Act of 2007, 42 USC 9801 *et seq.*, § 641 (2007).

17. Mead, S. (2014). *Renewing Head Start's promise: Invest in what works for disadvantaged preschoolers*. Retrieved from Bellwether Education Partners website: http://bellwethereducation.org/sites/default/files/Bellwether_Head-Start_July2014.pdf

18. Molina-Powell, K. (2014). 2013 Designation Renewal System: Analysis of the winners list. *Shine Early Learning Spotlight Newsletter, 1.3.* http://us2.campaign-archive1.com/?97af71a7649bfa2fdcccd1&id=c577635f52&e=b2d02a75a7

19. Lieberman, A. (2014, August 21). The results are in: Head Start re-competition, Round 2. Retrieved from New America Foundation website: https://www.newamerica.org/education-policy/edcentral/results-head-start-re-competition-round-two/

20. Office of Head Start, Early Childhood Learning and Knowledge Center. (2016). Review reports. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/grants/monitoring/review-reports.html

21. Mead, S., Mitchel, A. L., & Rotherham, A. J. (2015). *The state of the charter school movement.* Retrieved from Bellweather Education Partners website: http://bellwethereducation.org/sites/default/files/Charter%20Research%200908%20FINAL.pdf

22. U.S. Department of Health and Human Services, Administration for Children and Families. (2011). Part 1307: Policies and procedures for designation renewal of Head Start and Early Head Start grantees: § 1307.3. Basis for determining whether a Head Start agency will be subject to an open competition. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/standards/hspps/1307#1307.3.

23. U.S. Department of Health and Human Services, Administration for Children and Families. (2011). Head Start Program Performance Standards and other regulations: 45 CFR 1301–1311. Available from http://eclkc.ohs.acf.hhs.gov/hslc/standards/hspps

24. U.S. Department of Health and Human Services, Administration for Children and Families. (2016). Head Start Program Performance Standards: 45 CFR 1302.21(c)(2). Available from https://eclkc.ohs.acf.hhs.gov/hslc/hs/docs/hspss-final.pdf.

25. Robin, K. B, Frede, E. C., & Barnett, W. S. (2006). *Is more better? The effects of full-day vs half-day preschool on early school achievement* [National Institute for Early Education Research Working Paper]. Retrieved from http://nieer.org/publications/nieer-working-papers/more-better-effects-full-day-vs-half-day-preschool-early-school

26. Reynolds, A. J., Richardson, B. A., Hayakawa, M., Lease, E. M., Warner-Richter, M., Englund, M. M., . . . Sullivan, M. (2014). Association of a full-day vs part-day preschool intervention with school readiness, attendance, and parent involvement. *Journal of the American Medical Association, 312,* 2126–2134.

27. U.S. Department of Health and Human Services, Administration for Children and Families. (2011). Head Start Program Performance Standards and other regulations: 1302.31. Requirements of alternative agency. http://eclkc.ohs.acf.hhs.gov/hslc/standards/hspps/1302/1302.31%20Requirements%20of%20alternative%20agency..htm

28. U.S. Department of Health and Human Services, Administration for Children and Families. (2016). Head Start Program Performance Standards: 45 CFR 1302.51(b). Available from https://eclkc.ohs.acf.hhs.gov/hslc/hs/docs/hspss-final.pdf.

29. Pianta, R. C, Mashburn, A., Downer, J., Hamre, B., & Justice, L. (2008). Effects of web-mediated professional development resources on teacher-child interactions in pre-kindergarten classrooms. *Early Childhood Research Quarterly, 23,* 431–451.

30. Hamre, B., Pianta, R., Mashburn, A., & Downer, J. (2012). Promoting young children's social competence through the preschool PATHS curriculum and MyTeachingPartner professional development resources. *Early Education and Development, 23,* 809–832.

31. Mead, S., & Mitchel, A. L. (2016). *Moneyball for Head Start: Using data, evidence, and evaluation to improve outcomes for children and families.* Retrieved from Bellwether Education Partners website: http://bellwethereducation.org/sites/default/files/MoneyballforHeadStartFINAL.pdf

32. Improving Head Start for School Readiness Act of 2007, 42 USC 9801 *et seq.*, § 641A(a)(2)(C)(ii) (2007).

33. U.S. Department of Health and Human Services, Administration for Children and Families. (2014). *Head Start program facts fiscal year 2014.* Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/docs/hs-program-fact-sheet-2014.pdf

34. Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., & Squires, J. H. (2016). *The state of preschool 2015.* New Brunswick, NJ: National Institute for Early Education Research. http://nieer.org/research/state-preschool-2015

35. Minervino, J., & Pianta, R. (2014). Early learning: The new fact base and cost sustainability. In J. Minervino, *Lessons from research and the classroom: Implementing high-quality pre-K that makes a difference for young children.* Retrieved from Bill & Melinda Gates Foundation website: https://docs.gatesfoundation.org/documents/Lessons%20from%20Research%20and%20the%20Classroom_September%202014.pdf

36. Consolidated Appropriations Act of 2016, Pub. L. No. 114-113, Division H, Title II. (2015). Retrieved from https://www.congress.gov/bill/114th-congress/house-bill/2029/text?format=txt

37. Camilli, G., Vargas, S., Ryan, S., & Barnett, W. S. (2010). Meta-analysis of the effects of early education interventions on cognitive and social development. *Teachers College Record, 112,* 579–620.

38. Puma, M., Bell, S., Cook, R., Heid, C., Broene, P, Jenkins, F., . . . Downer, J. (2012). *Third grade follow-up to the Head Start Impact Study: Final report* (OPRE Report 2012-45). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Home visiting programs:
Four evidence-based lessons for policymakers

Cynthia Osborne

Summary. Home visiting programs (HVPs) aim to help low-income parents enhance their parenting skills and improve a host of early health and developmental outcomes for young children. Over the past five decades, numerous HVP models have been developed and implemented, albeit with modest or even null results, according to meta-analyses and comprehensive reviews. In 2010, in an effort to advance HVPs' effectiveness, federal lawmakers vastly expanded funding for HVPs with certain caveats, one being the requirement that the majority of programs be evidence based. Although the new requirement is a policy win, this review presents four main areas that must be addressed and improved upon if this new funding effort is to maximize positive outcomes. Pointedly, HVPs should have built-in flexibility for states to match the specific or unique needs of a family to a program model that has demonstrated effectiveness in meeting those specific needs. Further, program developers should clearly demonstrate what it is specifically about their model that works, in what context, and for whom. Ultimately, not unlike personalized medicine, state policymakers should target delivery of the right HVP model to the right family at the right time.

Home visiting is a promising early intervention strategy that aims to improve child and family outcomes by providing support, education, and access to resources for expectant parents and families with young children. Over the past five decades, numerous home visiting program (HVP) models have been developed with goals such as reducing child abuse and neglect, promoting healthy birth outcomes, increasing school readiness, and enhancing family economic self-sufficiency. Several HVP models have undergone rigorous evaluations to quantify the short- and longer term benefits for mothers and their children, and this evidence base has generated widespread hope that home visiting will reduce disparities in children's outcomes.

In 2010, Congress and President Obama established the Maternal, Infant, and Early Childhood Home Visiting Program (MIECHV), enacted as part of the Affordable Care Act and funded to the tune of $1.5 billion in formula grant funding over 5 years. The MIECHV

initiative was a massive scale-up of the Evidence-Based Home Visiting program of 2008 launched under President Bush. In an effort to improve outcomes achieved from HVPs, MIECHV requires states to spend at least three-quarters of the federal funds allocated on HVP models that meet the federally established criteria of evidence-based effectiveness.[1] To be considered evidence based, the HVP model must have been evaluated using a randomized control or quasi-experimental study design.

The decision of the federal government to rely on social science evidence to guide funding was hailed as a victory for both fiscal responsibility and evidence-based policy.[2] By spring 2016, 19 HVP models had been determined to meet the federal criteria and are on the list of approved programs from which states can choose.[3]

Demonstrating impact in randomized control trials, however, does not always translate to impact at the community level. The developers of the HVP models that are being used widely across the United States have

a responsibility to taxpayers and to the states to demonstrate that their models' effects can be retained when taken to scale. Scaling up any intervention is difficult, but unless programs retain their effectiveness when implemented widely, evidence-based programs will not fulfill their promise and policymakers may reduce or eliminate spending on home visiting.

Since 2011, I have been the lead evaluator of the Texas Home Visiting (THV) program, the largest HVP program in the country, serving over 6,500 families in 13 diverse communities across the state. The THV program has used four of the most common home visiting program models—Nurse-Family Partnership (NFP), Parents as Teachers (PAT), Early Head Start–Home Based (EHS-HB), and Home Instruction for the Parents of Preschool Youngsters (HIPPY)—although in 2015, we stopped using the EHS-HB. (See Table 1 for an overview of effectiveness of these programs—as well as another popular HVP—on six federal outcome measures.) Some communities implemented all four program models,

Table 1. Home Visiting Program Model Impacts on Federal Priority Outcome Domains

Outcome measure	Early Head Start – Home Based (EHS-HB)	Healthy Families America (HFA)	Home Instruction for Parents of Preschool Youngsters (HIPPY)	Nurse Family Partnership (NFP)	Parents as Teachers (PAT)
Maternal and newborn health	No effect	Favorable (secondary)	Not measured	Favorable (primary)	No effect
Prevention of child injuries, child abuse, neglect, or maltreatment and reduction of emergency department visits	Favorable (secondary)	Favorable (primary)	Not measured	Favorable (primary)	Favorable (primary)
Improvement in school readiness and achievement	Favorable (primary)	Favorable (primary)	Favorable (primary)	Favorable (primary)	Favorable (primary)
Reduction in crime or domestic violence	Not measured	Favorable (secondary)	Not measured	Favorable (secondary)	Not measured
Improvements in family economic self-sufficiency	Favorable (secondary)	Favorable (secondary)	Not measured	Favorable (primary)	Favorable (primary)
Improvements in the coordination and referrals for other community resources and supports	Favorable (secondary)	Favorable (primary)	Not measured	No effect	Not measured

Note. Source: US Department of Health and Human Services (2015). Home Visiting Evidence of Effectiveness (HomVee).

Primary outcomes are measured through direct observation, direct assessment, or administrative data or are self-reported data collected using a standardized (normed) instrument. **Secondary outcomes** include most self-reported data, excluding self-reports based on a standardized (normed) instrument. Data are accessed and adapted from "Home Visiting Program Model Effects" [Table], U.S. Department of Health and Human Services, Administration for Children and Families, 2015, http://homvee.acf.hhs.gov/EvidenceOverview.aspx?rid=4.

whereas others implemented only two, for a total of 34 MIECHV-funded programs across the state.

As part of the THV evaluation, my research team and I have visited each of the communities multiple times, repeatedly interviewed the program administrators and home visitors, conducted seven focus groups with parents, executed longitudinal surveys of the mothers (N = 1,698) and home visitors (N = 135), and analyzed data collected by the state from each of the four HVP models. This analysis, combined with two extensive evidence reviews that I completed with my staff,[4,5] illustrates the strengths and potential limitations of the evidence-based approach to home visiting. Texas constitutes a living laboratory that presents essential lessons for the future. Four of these lessons are especially pertinent; failing to heed them could jeopardize the success of the MIECHV initiative and evidence-based policymaking across the United States.

Lesson 1: Align the Strengths of the HVP Models with Community Goals

Within MIECHV, administrators often assume that because an HVP model is on the list of 19 federally approved programs, it will solve all family and early childhood problems. Administrators are rarely steeped in the home visiting evidence base and therefore may choose HVP models that are less than ideal for addressing the problems they are trying to resolve in either an individual family or the community as a whole.

No HVP model can do it all. There is no program model that has demonstrated improvement for each of the federal priority outcome areas stipulated in the MIECHV legislation. Too often, communities make the specious assumption that any HVP model will work for all populations and on any outcome. But program models vary considerably across a range of factors, including their goals, their target population, the curriculum, the required qualifications of home visitors, and the frequency and duration of the visits. The four program models used in Texas illustrate this variety.

Parents as Teachers (PAT) aims to provide general parenting education and serves a broad range of families, including pregnant women and families with children from birth through age 5 years. For higher risk families, the home visitor comes twice a month, although the standard program requires only a single visit per month.

Early Head Start – Home Based (EHS-HB), which uses the PAT curriculum in THV, serves low-income pregnant women and families with children from birth to age 3 years. The home visits are provided weekly, and the parents participate in several additional enrichment activities.

Home Instruction for the Parents of Preschool Youngsters (HIPPY) focuses on school readiness. The program does not have an income eligibility requirement and serves parents of children ages 3 to 5 years old. The program lasts 10 months and includes 30 weekly visits, plus group meetings.

Nurse Family Partnership (NFP) emphasizes maternal and child health; it has a higher recommended frequency and duration of visits than the other models and strenuous eligibility requirements. Recipients must be low-income, first-time mothers who are not more than 28 weeks pregnant. It is also the only model of the four used in Texas that requires the home visitor to have a bachelor's degree in nursing. The other models employ paraprofessionals or former program recipients to deliver the home visits.

This variation in service delivery and goals is mirrored in the outcomes for the models involved. All HVPs have met the evidence-based outcomes criteria on at least one of the six federally defined priority outcome areas: maternal health, child maltreatment, school readiness, crime or domestic violence, economic self-sufficiency, and referrals to services. But some models demonstrate impacts on multiple outcomes. Specifically, across the six benchmark areas prioritized by MIECHV, NFP shows at least one favorable impact in five areas; EHS-HB and PAT show favorable impacts in four and three of the six benchmarks, respectively; and HIPPY shows an impact in only one.

The reality is that no program model has proven benefits in all six federal benchmark outcome areas. Given that MIECHV requires states to demonstrate progress in four of the six priority benchmark areas, states may be wise to use several program models and models with more comprehensive impacts, which is the strategy followed by most states.

Often, a community chooses its HVP models on the basis of factors such as the age of the children served and whether a given model already exists within that community. Aligning the chosen model with particular community goals happens far too infrequently. In light of this, the federal government should require

that states and communities demonstrate their knowledge of the evidence base associated with their chosen program models and align HVP models with the specific outcomes the community is trying to achieve.

Lesson 2: Set Realistic Expectations

Communities often set unrealistic expectations for the programs they decide to use. All models have shown benefits on one or more outcomes in previous rigorous research, but the impacts are typically small, and they may not translate into large, community-level improvements. Meta-analyses and comprehensive reviews of home visiting evaluations find that most high-quality studies report null effects; even when effects are positive, the impacts are usually modest. In addition, the effects tend to be more pronounced among the most disadvantaged or high-risk subgroups.[6,7,8]

The attention home visiting receives in the media and from policymakers does not reflect the tepid impacts found in the evidence base. Indeed, President Obama's Plan for Early Education for All Americans cites evidence-based home visiting programs as having "been critical in improving maternal and child health outcomes in the early years, leaving long-lasting, positive impacts on parenting skills; children's cognitive, language, and social-emotional development; and school readiness."[9] In a similar vein, Nicholas Kristof, a *New York Times* columnist, commented in an op-ed coauthored with his wife Sheryl WuDunn that "the visits have been studied extensively through randomized controlled trials—the gold standard of evidence—and are stunningly effective."[10]

Home visiting programs are the most promising early childhood intervention we have, but they are not a magic bullet. A public dialog that sets realistic expectations for what home visiting programs can do for disadvantaged families and children will help states and communities understand whether their efforts are successful and aligned with reality. This will also help to avoid disappointment if future impacts continue to be null or modest.

To demonstrate the range of benefits—and the limits—found in home visiting, my colleagues and I examined findings reported in the literature for four widely used program models: EHS-HB, NFP, PAT, and Healthy Families America (HFA), a model commonly used in MIECHV-funded states that was designed to reduce child maltreatment. We examined a sample of important parenting behaviors, including prenatal care, breastfeeding, well-child visits and immunizations, learning support, and child maltreatment. We found that the HVP models generally have a robust impact on learning support and child maltreatment but limited or null impacts on the other parenting outcomes we examined.

For example, NFP is the only program model of the four that has demonstrated any impact on prenatal care at all. That beneficial outcome was measured through a study of NFP conducted in Elmira, New York, in the 1970s: Researchers demonstrated that mothers visited by program nurses were more likely to attend a childbirth class during pregnancy and knew more about available prenatal services. But even there, the program did not increase the level of prenatal care received.[11] The overwhelming majority of mothers who participated in EHS-HB and HFA received prenatal care services during their pregnancy, but we lack information on the comparison groups' outcomes, making it impossible to determine if the program had an impact.[12,13] Despite the fact that PAT serves mothers prenatally, researchers have not tested its impact on prenatal care.

The models have also had minimal and varied impacts on breastfeeding. Neither EHS-HB nor HFA demonstrated an impact on breastfeeding, and PAT's impact on breastfeeding has not been tested at all. NFP did demonstrate a positive impact on breastfeeding, but the findings were limited primarily to first-time African American mothers in Memphis in the early 1990s, 26% of whom initiated breastfeeding, compared with only 16% of mothers in the control group. At the 6-month follow-up, there was no difference between the groups in breastfeeding duration.[14] The impact on breastfeeding was not replicated in other NFP evaluations. Similarly, the four models also have had limited and varied impacts on well-child visits and immunizations.[15]

The evidence for impacts on learning support is stronger than the evidence for the other outcomes. Indeed, EHS-HB, NFP, and PAT all show positive impacts on parent's support for children's learning, although the construct was measured differently across programs. For example, EHS-HB[16] and NFP[14] showed positive impacts on the Home Observation Measurement of the Environment (HOME) Inventory, which measures the quality and quantity of stimulation and support available to a child in the home environment. The results for NFP

applied to the mostly African American sample of high-risk mothers in Memphis and the most disadvantaged mothers in Elmira,[17] but marginal results were found for the more diverse sample of mothers in a Denver study.[18] EHS-HB also showed modest impacts on reading to children daily by the time they reached kindergarten. And PAT showed a positive impact on reading aloud and parent engagement, but the findings were limited to the most disadvantaged children in the study.[19] The HFA studies found virtually no impact on learning supports. Each of the four models had a positive impact on reducing child maltreatment, but the findings were more robust among the most disadvantaged groups.

The HVP models have demonstrated impacts on several outcomes not discussed here, but this brief summary sheds light on the mixed and generally modest results found in the evaluations that make up the evidence base. States and communities should not rely on HVPs alone to reduce childhood adversity and create better outcomes for children and families. Home visiting programs should be one component of a continuum of care that supports parents and children. To gain the most benefit, communities and administrators must understand the impacts they can expect from each well-implemented HVP model.[20] They would also do well to remember that impacts at the population level tend to be far more difficult to detect than impacts at an individual level, particularly if the individual changes are modest. To prevent disillusionment, politicians, policymakers, pundits, and academics should set fair expectations for HVP results rather than engage in hyperbole.

Lesson 3: Understand Why Each HVP Model Works and In What Context

The third important lesson from the evidence-based home visiting approach is that context matters. It is unreasonable to expect that the results found in the evidence base will be replicated precisely in the real world. Indeed, the home visiting evidence base is replete with mixed results across models and, more important, within each model. The findings from one HVP model study are seldom replicated when the model is implemented in a different geographic or demographic context.[21]

This lack of replication and generalizability means that the home visiting evidence base is limited. Program model developers have only been required to demonstrate whether their program works; they have not had to illuminate what about the program model works best, under what circumstances, and for whom. Although this information is difficult to determine through large, rigorous impact evaluation studies, strong implementation studies and smaller outcome studies that examine various aspects of the program models can be valuable tools. Without this additional information, states and communities lack guidance on how to reap the benefits promised by the evidence-based model they are using.

An excellent example of the importance of context matters in HVP model impacts is demonstrated in studies of NFP, the longest running and most rigorously evaluated home visiting model. Over the years, three major studies of NFP, based in Elmira, Memphis, and Denver, respectively, have evaluated specific outcomes, yet none has been proven and replicated across all three studies. The Elmira study has so far demonstrated the largest and longest term impacts, whereas fewer impacts were found in Memphis or Denver. (The variation may be due, in part, to the different time periods in which the studies were conducted and the cultural and demographic differences in the populations studied. The Elmira study took place in 1978–1980 and included 400 first-time mothers; 90% of the sample was White. The Memphis study of 1,139 first-time mothers took place a decade later; there, 92% of participants were African American. The Denver study of 735 first-time mothers ran from 1994–1995 and had a more diverse sample: 46% were Hispanic, 36% were White, and 15% were African American.)

Also illustrative of this lack of replication/reproducibility in studies of HVPs: while five evaluations of HFA have been reviewed by the federal government, the results from one study have not typically been replicated in another context; also troubling is that the developers of the model have not provided enough insight as to why.

Given the inability to replicate findings from one context to another, states and communities cannot feel confident that benefits proven in one population or situation will work elsewhere. In Texas, this seems to be the case. One example of variation in Texas outcomes is with breastfeeding initiation: Among nine communities serving pregnant mothers using various HVPs, rates of breastfeeding initiation ranged between 19% and 95%. One program model had very high rates of initiation

(over 80%) in each community, whereas another model showed considerable variation across communities (from 19% to 41%).

Low adherence to model fidelity may be another reason for the large variation in outcomes.[22] Few home visitors pay strict attention to their model's curriculum. A common refrain from home visitors is that although they begin a home visit with the intention of addressing the prescribed topic for the visit, "life gets in the way," and they spend time meeting the individual needs of the mother. A discussion on car seat safety, for example, seems less important than helping a mother who is about to be evicted locate the resources she needs to retain her home. Allowing home visitors the flexibility to meet the mother's goals and needs is part of the philosophy of some of the programs, yet that makes it difficult to determine what information is actually being shared with parents consistently. Measuring fidelity to the curriculum and learning the core principles of the model are nearly impossible with this flexible approach.

Partial participation and attrition from the program offer additional explanations for variation in results. Each program is committed to serving its families, but programs that enroll teen parents or parents with high levels of risk have greater difficulty meeting with the parents as planned. Home visitors lament that missed appointments and families leaving the program before completion interfere with meeting a family's goals. To the extent that dosage and attrition differ across HVP models, outcomes are likely to be affected.

It is no surprise that outcomes will vary given the variation in inputs and contexts. But model developers need to better define what level of variation is part of the model and what variation conflicts with the model's fidelity. If delivery of the curriculum is believed to be what is responsible for the model's success, then the curricular elements need to be identified and replicated each time the model is implemented. Currently, neither the home visitor, the developers of the models, nor the states are closely monitoring fidelity because no one is certain what *fidelity* actually means. If fidelity to the model is not a priority and is not adhered to, then is MIECHV really an evidence-based policy approach?

To be on the federal government's approved list of evidence-based programs, the HVP model's developers should be responsible for identifying their core program components and activities. The developers should also explain what it is about their model that produces specific outcomes, as well as why, for whom, and under what conditions. Without this peek inside the black box, communities do not know which elements of the model to faithfully replicate and which elements they could alter to fit their circumstances.

Identifying core elements of HVPs may ultimately permit states and communities to move away from strict adherence to a particular program model and develop an approach that is tailored to the varying needs of families. Ideally, every mother during pregnancy and at birth would receive one home visit devoted to parenting education, screening for potential risk factors, and connecting with necessary resources. Mothers with identified risks would receive additional home visits commensurate with their needs. But without understanding the core elements of a model, it is virtually impossible to custom design a given program for each parent's needs.

Lesson 4: Innovation Is Important for Ongoing Success of Home Visiting

A final concern about the evidence-based policy approach is that it may fail to foster innovation. The existing evidence base must be strengthened through evaluations of new program models that meet families' needs. It is encouraging that MIECHV contains important research elements that may strengthen the evidence base. Specifically, MIECHV allows states to use one-quarter of formula funds on program models that have not yet met the high standard of evidence but are undergoing evaluations. (MIECHV also funded the Maternal and Infant Home Visiting Program Evaluation, which will examine the outcomes of the four most widely used program models—NFP, EHS-HB, HFA, and PAT—and study variation in program implementation.)[23]

In addition to improving the existing research base, researchers and policymakers need to develop new program models that address today's most pressing health and social policy issues. For example, prenatal smoking was a serious concern in the 1970s, and the NFP model demonstrated a substantial impact on reducing the incidence of prenatal smoking. However, today, maternal prenatal obesity and early childhood obesity are widely prevalent health concerns, yet no program has been designed to address obesity.

Innovation may also be stifled by inertia. Over time, communities accumulate knowledge and expertise on

how to administer and support a particular HVP model, and the community may be reluctant to change course, even if that model is not making a difference in families' outcomes. To correct for this, an evidence-based approach to policy making should entail continuous assessment and modification.

To strengthen the evidence base, the federal government must implement a mechanism in MIECHV that requires program model developers to continually evaluate and enhance their models to remain on the list of evidence-based models. Models yielding results that demonstrate that their impacts are robust across time and populations and models that identify their core elements should be considered evidence based. Other models might be considered promising practices.

Policymakers should move away from offering families pre-determined programs and move toward providing families with what they actually need. In practice, however, communities are not typically aligning a family's needs with a model designed to meet those needs. Rather, communities are generally delivering the model they offer to any family who meets the eligibility requirements, regardless of that family's needs. Some families may be overserved and others underserved in a quest to implement a preferred model(s). For example, a family may need 2 weeks of minimal services to connect them to other resources but, instead, the family is put into a program that offers services for years. Alternatively, a family may need intense case management, but the program model in which they are enrolled may provide only monthly home visits. With greater clarity over what it is about home visiting services that impacts family outcomes, programs could move toward offering more individualized services aligned to families' needs.

Conclusion

Using evidence to inform decisions about what programs to fund is a reasonable and prudent approach to policymaking. But taking an evidence-based policy-making approach to home visiting means that program developers and administrators must identify what it is about their HVP model or models that positively impacts families and which family needs are best met by their programs. Policymakers and administrators in states and communities have a responsibility to know the evidence base and implement programs that will address the goals they establish. It is imperative to align a model's evidence

of impacts with the needs it is being put in place to meet. Finally, federal policymakers should take steps to strengthen the evidence base and put it to targeted use while simultaneously fostering innovation. Using the evidence base to inform efforts to target services more effectively to families so that they receive the right level and elements of services to meet their needs and improve their children's lives is sound policy.

author affiliation

Osborne, LBJ School of Public Affairs, University of Texas at Austin. cosborne@prc.utexas.edu

References

1. U.S. Department of Health and Human Services, Administration for Children and Families. (2013). Home visiting evidence of effectiveness: About us. Executive summary. Retrieved from http://homvee.acf.hhs.gov/About-Us/5/Executive-Summary/20/2
2. Haskins, R., & Margolis, G. (2014). *Show me the evidence: Obama's fight for rigor and results in social policy.* Washington, DC: Brookings Institution Press.
3. U.S. Department of Health and Human Services, Administration for Children and Families. (n.d.). Home visiting evidence of effectiveness: Models. Retrieved July 30, 2016, from http://homvee.acf.hhs.gov/Models.aspx.
4. Osborne, C., Bobbitt, K., & Ansari, A. (2015). *From randomized controlled trials to community-level change: What should be expected when taking home visiting programs to scale?* [Working paper]. Austin: The University of Texas at Austin, Child and Family Research Partnership.
5. Bobbitt, K., & Osborne, C. (2015). *Reality check: Can the effects of home visiting programs found in RCTs be replicated in the real world?* [Working paper]. Austin: The University of Texas at Austin, Child and Family Research Partnership.
6. Astuto, J., & Allen, L. (2009). Home visitation and young children: An approach worth investing in? *Social Policy Report, 23*(4). Retrieved from http://srcd.org/sites/default/files/documents/23-4_astuto_allen.pdf
7. Azzi-Lessing, L. (2011). Home visitation programs: Critical issues and future directions. *Early Childhood Research Quarterly, 26,* 387–398.
8. Sweet, M., & Appelbaum, M. (2004). Is home visiting an effective strategy? A meta-analytic review of home visiting programs for families with young children. *Child Development, 75,* 1435–1456.
9. White House, Office of the Press Secretary. (2013, February 13). *Fact sheet President Obama's plan for early education for all Americans* [Fact sheet]. Retrieved from https://www.whitehouse.gov/the-press-office/2013/02/13/fact-sheet-president-obama-s-plan-early-education-all-americans
10. Kristoff, N., & WuDunn, S. (2014, September 12). The way to beat poverty. *The New York Times.* Retrieved from http://www.nytimes.com
11. Olds, D., Henderson, C., Tatelbaum, R., & Chamberlin, R. (1986). Improving the delivery of prenatal care and outcomes of pregnancy: A randomized trial of nurse home visitation. *Pediatrics, 77,* 16–28.

12. Chazan-Cohen, R., Raikes, H., & Vogel, C. (2013). Program subgroups: Patterns of impacts for home-based, center-based, and mixed-approach programs. *Monographs of the Society for Research in Child Development, 78,* 93–109.

13. Daro, D. A., & Harding, K. A. (1999). Healthy Families America: Using research to enhance practice. *The Future of Children, 9,* 152–176.

14. Kitzman, H., Olds, D. L., Henderson, C. R., Jr., Hanks, C., Cole, R., Tatelbaum, R., . . . Barnard, K. (1997). Effect of prenatal and infancy home visitation by nurses on pregnancy outcomes, childhood injuries, and repeated childbearing: A randomized controlled trial. *JAMA: The Journal of the American Medical Association, 278,* 644–652.

15. Landsverk, J., Carrilio, T., Connelly, C. D., Ganger, W., Slymen, D., Newton, R., et al. (2002). *Healthy Families San Diego clinical trial: Technical report.* The Stuart Foundation, California Wellness Foundation, State of California Department of Social Services: Office of Child Abuse Prevention. Referenced on the HomVee evidence website: http://homvee.acf. hhs.gov/Effects/1/Healthy-Families-America--HFA-/10/ Child-Health/2/4/#2.

16. Harden, B. J., Chazan-Cohen, R., Raikes, H., & Vogel, C. (2012). Early Head Start home visitation: The role of implementation in bolstering program benefits. *Journal of Community Psychology, 40,* 438–455.

17. Olds, D., Henderson, C., & Kitzman, H. (1994). Does prenatal and infancy nurse home visitation have enduring effects on qualities of parental caregiving and child health at 25 to 50 months of life? *Pediatrics, 93,* 89–98.

18. Olds, D., Robinson, J., Pettitt, L., Luckey, D. W., Holmberg, J., Ng, R. K., . . . Henderson, C. R., Jr. (2004). Effects of home visits by paraprofessionals and by nurses: Age 4 follow-up results of a randomized trial. *Pediatrics, 114,* 1560–1568.

19. Wagner, M., Spiker, D., & Linn, M. (2002).The effectiveness of the Parents as Teachers program with low-income parents and children. *Topics in Early Childhood Special Education, 22,* 67–81.

20. Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science, 16,* 893–926.

21. Azzi-Lessing, L. (2013). Serving highly vulnerable families in home-visitation programs. *Infant Mental Health Journal, 34,* 376–390.

22. Paulsell, D., Del Grosso, P., & Supplee, L. (2014). Supporting replication and scale-up of evidence-based home visiting programs: Assessing the implementation knowledge base. *American Journal of Public Health, 104,* 1624–1632.

23. Michalopoulos, C., Lee, H., Duggan, A., Lundquist, E., Tso, A., Crowne, S. S., . . . Knox, V. (2015). *The Mother and Infant Home Visiting Program Evaluation: Early findings on the Maternal, Infant, and Early Childhood Home Visiting Program. A report to Congress* (OPRE Report 2015-11). Washington, DC: U.S. Department of Health and Human Services.

Launching Preschool 2.0: A road map to high-quality public programs at scale

Christina Weiland

Summary. Head Start and other publicly funded preschool programs are some of the most popular government programs in the United States, and in recent years officials have explored expanding public preschool and making it universal. However, several recent large-scale studies have raised questions about the benefits of these programs for participants and for society, as well as whether high-quality preschool is achievable on a large scale. This article reviews the available evidence on these questions and also what is known about the quality of various types of existing programs. The evidence indicates that the curriculum and professional development choices of most programs are out of step with the science of early childhood education and that this has made preschool programs less effective than they could be. The Boston Public Schools prekindergarten program can be used as a case study in better practice preschool implementation. Evaluation of this program shows that high-quality public preschool is achievable on a large scale if localities make the right investment and implementation decisions.

Over the last 3 years, publicly funded preschool has been in the policy spotlight. At the federal level, President Obama called for universal access to high-quality preschool for 4-year-old children in his 2013 State of the Union address. Several preschool proposals at the federal level soon followed. Not content to wait for federal action, states and cities around the country have introduced new public preschool programs or expanded existing ones. Preschool is also popular with the public. A recent survey found that 86% of Americans thought the federal government should help states and local communities build better preschool services and improve access.[1] Preschool, to quote one of the many news articles that have covered it since 2013, is "having its moment."[2]

This spike of interest in public preschool is due to a confluence of factors. Science has been one driver. Research across neuroscience, economics, and developmental psychology has converged on the importance of stimulating early childhood experiences and relationships in creating a foundation for lifelong success and on the potential for high-quality early education to

provide this foundation. Family needs are another driver. Large increases in maternal employment over the last several decades, especially among lower income families, have led to increases in rates of out-of-school care early in life.[3] At the same time, preschool is not cheap. At a time when average hourly wages have stagnated or fallen,[4] full-time preschool for a 4-year-old costs an average of $8,000 in the United States,[5] or 15% of the nation's average pre-tax family income[6] and more than 25% of earnings for most families in the lowest two income quintiles. Within a national conversation on rising income inequality, expanded support for public preschool has emerged as one that matches how Americans prefer to give—via provision of direct services.[7]

As preschool has come to the forefront, a healthy debate on its merits has emerged.[8,9] The debate has focused on two sets of questions. The first set of questions asks, essentially, Is public preschool worth the investment? That is, does preschool improve children's school readiness? Do its benefits last? If so, for how long? Let's call these *Preschool 1.0 questions*.

The second set, or *Preschool 2.0 questions*, focuses on the particulars of program models. Is high-quality preschool scalable? What specific program elements work best at ensuring high quality and promoting strong and lasting gains for children?

Existing research provides an extensive, although imperfect, evidence base for addressing both sets of questions. In this article, I summarize this evidence base. I give particular attention to Preschool 2.0 questions and debate for several reasons. First, regarding Preschool 1.0 questions, given working families' demand and needs, publicly funded preschool is unlikely to go away. 42 states and a handful of cities already have funded their own public preschool programs,[10] and the federal Head Start program serves nearly one million preschoolers.[11] Although further expansion is a high policy interest area and there are sizable access gaps by family income,[12] Preschool 2.0 questions are arguably more cross-cutting for policy because they apply to both existing and future programs. Third, the full evidence base on Preschool 2.0 questions is generally less well-known than the Preschool 1.0 question evidence. Most programs are not implementing evidence-based models, and instructional quality in particular is low.[13] To date, policymakers have not insisted that public preschool programs adhere to the evidence base that details the practices most likely to produce positive

outcomes for children. Using the example of the Boston Public Schools prekindergarten program, where I have been a lead research partner for the past 7 years, I offer a case study of what it takes to implement evidence-based preschool models on a large scale. I conclude by identifying critical areas for new research and discussing the role of policy in raising preschool quality.

Preschool 1.0 Question: "Should we expand"?

Preschool 2.0 Question: "What should we expand and how do we make existing prekindergarten programs better?"

The Preschool 1.0 Evidence Base

There is a large evidence base on the benefits of preschool for participants.[14–17] In the short term, a recent comprehensive meta-analysis project that analyzed results from 84 rigorous evaluations of preschool programs conducted since 1960 found that, across the great variety of program models, locations, and samples, preschool boosted children's kindergarten readiness by an effect size of a 0.21 standard deviation (SD) for cognitive and achievement skills.[18] (The 0.21 SD estimate is weighted for precision. The unweighted estimate is 0.35 SD.) An effect size is a statistical yardstick that permits comparison of program effects across settings and measures. Researchers tend to classify a small effect as less than 0.4 SD, a medium effect as 0.4 to 0.6 SD, and a large effect as greater than 0.6 SD.

Research on the effects of preschool on children's socioemotional development (that is, positive behaviors showing empathy, cooperation, or prosocial orientations, or problem behaviors such as antisocial, aggressive, hyperactive, impulsive, withdrawn, depressed, or anxious behaviors) has been sparse and the pattern of findings has been somewhat inconsistent.[17] A recent meta-analytic study helped reconcile this evidence. The study found that when improving socioemotional skills is an explicit program goal, there are small, consistent improvements in those skills for children.[19]

In the medium term, preschool has shown small to moderate effects in reducing grade retention and special education placement in the kindergarten through 12th grade (K–12) years. A recent meta-analysis found average reductions of 0.29 SD or 10.1

percentage points for grade retention and 0.40 SD or 12.5 percentage points for special education placement.[20] Cognitive gains from preschool tend to mostly or entirely fade out by the end of third grade; recent work suggests most of the eventual total decline occurs within 1 to 2 years after preschool.[21] The mechanisms behind cognitive skill fade-out (or control group catch-up) and the program features that best prevent it are not well understood. Some work suggests the preschool boost is more likely to be sustained if participants have high-quality elementary school experiences.[22,23] Another study suggests the effects of an early childhood mathematics curriculum lasted into early elementary school only if kindergarten instruction was aligned with preschool instruction.[24] It is also possible that having more peers who attended preschool might sustain effects by allowing teachers to teach to a higher skill level rather than focusing on catching up preschool nonattenders.[25]

In the longer term, rigorous studies that have followed preschool participants into adulthood have found a host of long-term benefits, such as increases in college enrollment, decreases in incarceration rates, and decreases in teen pregnancy.[26–29] Despite these positive results, these studies may represent a biased sample of all of the studies conducted because studies that fail to show preschool's benefits may have been rejected for publication or never submitted in the first place. Studies that fail to generate positive results suffer this fate often enough that social scientists have a name for it: the "file-drawer problem."

Despite that caveat, as others have pointed out,[30] all three available national data sets for studying long-term benefits of Head Start have produced evidence of long-term benefits for participants. Cost–benefit work based on the Perry Preschool, Abecedarian, and Chicago Child-Parent Centers evaluations has suggested that preschool has a robust return on initial investment in the long term, with estimates as high as 10%.[31–33] There have been at least three "back of the envelope" approximations for Head Start,[30,34,35] all of which suggest it passes a cost–benefit test.

To estimate effects of preschool on adults, researchers have to wait until preschoolers reach adulthood. Thus, the context of these longer term studies is very different from the context for today's preschools. Today, more children attend nonparental care than did in the past, changing the condition against which

preschool is evaluated.[17] Parenting has also changed: Parents today invest more time and money in their children's learning, on average, than did previous generations.[36,37] A study of the current-day Tulsa program's likely cost–benefit ratio, using a projection method to extrapolate students' future earnings from their kindergarten test scores, suggests robust returns that are not unlike those of some older studies—$3 returned per $1 spent.[38] However, the applicability of the results of older, longer term studies to today's policy decisions and the specific drivers behind these effects are open questions. It is not yet known if today's preschool programs will yield benefits to participants and society similar to the benefits provided by programs from earlier decades.

The Preschool 2.0 Evidence Base

What specific program elements work best at ensuring high quality and in improving children's kindergarten readiness? The most important evidence comes from across-study differences in program quality and from studies that have randomly assigned children to different preschool program elements (as compared with business-as-usual preschool). I begin by defining *preschool quality* and then delve into this evidence base.

What Is Preschool Quality?

High-quality preschool tends to be conceptualized in two buckets: structural factors and process factors. *Structural quality* emphasizes broad characteristics of the preschool setting like teacher-to-student ratios, total class size, teacher education and training, and the safety of the classroom for young children. *Process quality* refers to the nature of the interactions between teachers and children and between the children themselves and the richness of specific learning opportunities within the classroom. Given typical input levels in the United States today, structural quality sets the stage for process quality to occur, but it does not guarantee that it will.[39]

Across-Study Patterns

In practice, measuring quality has been challenging (see the online Supplemental Material for more details). Nonetheless, the existing evidence base on the average effects of a given preschool program versus local alternatives indicates that programs with higher instructional

quality tend to have initially larger benefits for participants, and these larger initial benefits tend to yield larger lasting effects into adulthood, as compared with lower quality programs that have initially small effects.[27,30,40] (Factors other than program quality are also drivers of the size of preschool impacts on participants, particularly child characteristics and children's likely care settings in the absence of prekindergarten. I focus on program quality for space reasons and because it arguably is the most cross-cutting factor in current policy debates.)

Recent studies have mirrored this pattern of larger, more lasting benefits in programs whose initial benefits for participants are larger. Studies of the benefits to participants in Head Start and the Voluntary Pre-K for Tennessee program found that children with 1 year of preschool had small cognitive and social-emotional benefits. By the end of first grade, nonparticipants in these programs largely caught up to preschool attendees in their cognitive, academic, and socioemotional skills. At the end of third grade, there were no benefits to Head Start participants.[41] The Tennessee study recently found evidence of negative impacts for participants on a summative cognitive measure and on children's mathematics scores.[42] In contrast, the Tulsa prekindergarten program had initially large benefits on children's literacy skills, moderate benefits on children's mathematics skills, small benefits in reducing children's timidity and increasing their attentiveness, and no effects on children's aggressive or hyperactive behavior.[43,44] The initial boost in Tulsa on math of 0.38 SD lasted through the end of third grade (0.18 SD) for a cohort that experienced a mature and presumably higher quality version of the program.[45] Effects on reading did not persist for either Tulsa cohort.

Peeks inside the classrooms of these programs suggest these disparate findings may be due in part to differentials in quality. Data suggest that Head Start structural[42] quality and emotional quality are good but that instructional quality is inadequate.[46] In Tulsa, preschool instructional quality was approximately 0.33 SD higher than current Head Start levels[47] and 0.97 SD higher than Head Start average levels in 2009.[46,48] The Tennessee study did not use the same instructional quality measure as Tulsa and Head Start, which limits direct comparison of quality levels. However, in a statewide representative study of the Tennessee program, 85% of classrooms did not meet the "good"

benchmark for overall quality. Detailed time-use data collected in these classrooms revealed that children spent more of the school day in whole group instruction (32%) than in more learning-efficacious small-group settings (24%). Children spent 44% of the day in noninstructional activities.[49]

Curriculum and Professional Development Research

Rigorous studies of curricula and professional development over the last decade have found that some approaches work better than others in improving preschool quality and child outcomes. Across diverse contexts and curricula, the strongest route to attaining program quality appears to be coaching by an expert mentor, paired with a domain-specific curriculum that has a specified scope and sequence.[39] Having a mentor who observes teachers' in-classroom work with students, troubleshoots problems in teacher practice, and supports teachers' curricular implementation is more effective in changing teacher practices than the usual approach of professional development, which is one-shot workshops. Curricula that support children's natural developmental trajectories via a specific scope and sequence for specific skills appear to be more successful in promoting children's gains in prekindergarten than those that focus on the whole child, without a specified scope and sequence.[39] Pairing curriculum and coaching seems to be key. At least one test of coaching without a sequenced curriculum showed improvements in quality, but those improvements were not sufficient to produce impacts on children's cognitive outcomes.[50]

Some studies have tested implementing more than one domain-specific curricula supported by coaching and training and found evidence of positive effects on targeted child outcomes.[51,52] This is important because preschool programs and teachers are charged with improving children's readiness for school across multiple domains—for example, language, literacy, mathematics, and socioemotional skills, not just literacy skills. Further, children from lower income families lag substantially behind their more advantaged peers across multiple developmental domains,[53] and early skills across multiple domains predict later development.[54] It is interesting that some curricula do have cross-domain effects.[55,56] The math-focused Building Blocks curriculum, for example, emphasizes having children express

their mathematical ideas and thinking through language. It has shown positive impacts on children's executive function skills and on measures of oral language such as use of complex utterances.

Existing Public Preschool Programs: Models and Quality Levels

The evidence base on preschool quality drivers is still emerging, and very few localities have invested in evidence-based, domain-specific curriculum and coaching supports. Rather, the most popular curricular choices in programs nationally are comprehensive, whole-child-focused curricula.[,13, 57] Some of these curricula show pre–post gains for enrolled children in descriptive research, and some descriptive research suggests some of these curricula may outperform others.[58] But the rigorous work that exists suggests that they have at best limited evidence of effectiveness in improving children's kindergarten readiness.[59]

The What Works Clearinghouse, which reviews rigorous studies of preschool curricula and gives each curriculum an effectiveness rating, rates one of these common choices—Creative Curriculum—as having an effectiveness rating of zero for children's mathematics, oral language, phonological processing, and print knowledge skills.[60] Yet, on a recent list of approved curricula for state-funded preschool programs, Creative Curriculum was the curriculum most frequently mentioned.[61] It was also the most commonly used curriculum in Head Start in 2009[57] and the second most commonly used curriculum in an 11-state study of prekindergarten.[13]

As for professional development models, data on large-scale systems are neither available nor tracked. Traditionally, teachers are supported through training, often with a specific number of hours required per year.

There has been no published research on how program model decisions are made across localities. One possible reason for the choice of comprehensive curricula maybe the requirement in many places to have curricula that cover all child developmental domains—the historical whole-child focus in early childhood.[59] Another reason may be program requirements for teachers and programs to collect data on children's progress, because some comprehensive curricula are paired with specific collection tools and, in 2014, curriculum-paired tools were among the most

commonly approved for fulfilling child assessment requirements.[61] In addition, simple familiarity and history may play a role. Domain-specific curricula are generally newer than comprehensive curricula, meaning staff who make curriculum decisions are less likely to have experience with them. Today, no preschool policy requires the use of evidence-based curricula.

Regardless of the reasons behind local programmatic decisions, the net result of current choices—of which curricula and professional development are just two— appears to be mixed in terms of the quality experienced by enrolled children. Structural quality and emotional climate are at levels considered good by experts.[4] Programs score poorly, however, on instructional quality—the aspect of quality that appears to matter most for children's cognitive gains in preschool.[62]

Improving Quality and Outcomes at Scale: The Boston Prekindergarten Program

The Boston Public Schools prekindergarten program, which I have investigated with colleagues since 2007, is a notable exception in its curricular and professional development choices and investments. In 2007–2008, Boston implemented two domain-specific curricula supported by biweekly coaching by expert, experienced former teachers across district prekindergarten classrooms. After 2 years of implementation, this program model produced meaningful impacts on targeted (language, literacy, mathematics, and socioemotional skills) and nontargeted skills (executive function).[51] Impacts on children's vocabulary and math skills were the largest across rigorous evaluations of the effects of large-scale, public prekindergarten programs. Effects were particularly pronounced for Hispanic students, dual-language learners, children from low-income families, and children with special needs.[51,63]

Although Boston is a single district, it is a large one, with historically stubborn achievement gaps. As a case study, its story offers several lessons for scaling high-quality preschool, which I describe below.

Structural Quality Investments Are Not Enough

From the program's beginning in 2005, Boston made strong investments in structural quality. Under policies unusual for prekindergarten, teachers were paid on the same scale and were subject to the same educational

requirements (state certification and master's degree within 5 years) as K–12 teachers. In practice, this meant that by 2008, the majority held master's degrees and were paid a base salary of approximately $60,000 per year, per the district's K–12 salary schedule. Classrooms were also staffed by a paraprofessional, bringing the maximum adult-to-child ratio to 1:11. The program was based entirely in public schools. The program was open to all children in Boston via a lottery assignment system. Consequently, many classrooms were mixed income, an approach we later found to be associated with stronger vocabulary and executive function gains for children in the program (versus low-income only)[64] There was no consistent program curriculum or coaching system.

Despite these unusually strong structural investments, researchers at the Wellesley Center for Women found that process quality—particularly instructional quality—was low in the first few years of the program.[65]

Implementing Curricula and
Improving Quality at Scale

Following the Wellesley report, Boston slowed down the planned pace of expansion and upped the district's investment in program quality. With consultation from experts, the Boston Public Schools Department of Early Childhood selected two domain-specific curricula already in place and working well in some district schools: the language and literacy curriculum Opening the World of Learning and the mathematics curriculum Building Blocks. The district created an integrated curricular guide for each unit, with concrete guidance on how to implement Opening the World of Learning and Building Blocks together. They also made sure teachers had all of the materials they needed to implement the curricula at high levels,[66] such as specific items for center time that reinforced the unit theme or specific vocabulary being taught.

Teachers received support in implementing these curricula from training sessions (13 days total, in the summer and across the school year) and from approximately biweekly visits from expert coaches. The coaches were free to focus on any areas where the teachers' practice needed support. For some, this could mean considerable attention to classroom management before focusing on the curricula. For strong teachers,

this could mean extending the curricula beyond what was written to go deeper into a group of children's particular interests.

In data collected 2 years after Boston's quality investments, Boston preschool classrooms showed the highest average instructional quality of a large-scale program to date: 1.7 to 2.4 SDs higher than current Head Start quality nationally. (The SD range is due to the choice of SD used in the comparison. Boston's CLASS [Classroom Assessment Scoring System] Instructional Support score average was 4.3 and its SD was 0.84,[67] whereas the analogous mean and SD for Head Start were 2.9 and 0.58, respectively.[47]) Quality data were used at the district level to pinpoint and target particular district weaknesses. Scores were not used to reward or punish teachers; they were used for program decisions and improvement only.

Research as a Tool for Program Building

Another hallmark of the Boston program has been its strong emphasis on data-driven decision making—for example, changing course on the basis of quality data, choosing strong curricula that resonate locally, and partnering with researchers both to study specific program elements and on internal efforts that informed specific decisions. Research in Boston has not been a tool just for grading the preschool's effectiveness but for building the program—an approach that continues in Boston's current efforts to align its preschool to grade 3 (P–3) curricula and professional development and in the expansion of its models into community-based centers in Boston.

Moving Forward: Research
and Policy Implications

Despite the considerable learning over the last few decades, there are critical areas of need for more research. More research is needed on the contributions of other potential active ingredients in preschool programs besides curriculum and professional development, including rigorous research that randomly assigns children to preschool classrooms with different peer conditions (for example, all low-income backgrounds versus a mix of income backgrounds) and that compares teachers paid on the same scale as

K–12 teachers with those paid considerably less. More work is needed on creating and testing new integrated, domain-specific curricula, such as the National Science Foundation–funded Connect4Learning curriculum, which was developed by experts in mathematics, language, literacy, science, and socioemotional development. There is also a need to integrate existing, separate domain-specific curricula to streamline logistical implementation barriers for programs. Also needed is more research on how program models operate differently under different auspices and trade-offs of one auspice versus another. As others have argued, more research is needed on which early skills are the most critical to develop in preschool[18] and on more effective ways to measure quality.[62]

Advancing the science of preschool education also necessitates research on preschool for 3-year-olds and on the after-preschool years. Increasingly, children enter preschool at age 3 years.[3] To date, there has been very little research on optimal experiences and sequencing of instruction for 2 years of preschool. Also needed is more research on the mechanisms of catch-up and fade-out, a need that stands to be partially filled by an upcoming investment in an Early Learning Network by the Institute of Education Sciences, which includes Boston as a study site. Rigorous tests of P–3 models to support children's early learning are needed, as no proven P–3 models currently exist.

However, in the current absence of answers to these and other important questions, existing public preschool programs and policymakers are faced with doing the best they can with today's evidence. At this juncture, on the basis of decades of research, it is known with great confidence that high-quality preschool increases children's school readiness. There are wide income-based disparities in access to quality prekindergarten programs nationally and in kindergarten readiness, gaps that expanded access to publicly funded preschool can stem.[3] Research from past models suggests benefits can last into adulthood. Given differences in today's context versus the contexts of the older studies, there is no ironclad guarantee that expanding access to high-quality public preschool today will benefit participants and society in the long run. Results so far are consistent with those of previous studies—universal fade-out of the boost on cognitive test scores in K–12, more lasting

effects in the medium term in higher quality versus lower quality programs, and a robust projected return on investment.

Also known is that most public preschool programs are not making programmatic decisions that match the current science. Here, policy can play an important role in requiring programs to choose proven curricula and professional development models and to update their decisions to align with new findings. Also needed are new national quality standards that track additional proven and potential active ingredients, including which curricula are used, with what specific supports for teachers; whether preschool teachers are held to the same standards as K–12 teachers and whether they receive the same pay; and the socioeconomic composition of children's peers. Ultimately, the potential for preschool to improve the life prospects for young children depends on following the evidence not just on whether society should invest in preschool but regarding the more neglected question of how to implement specific program elements at scale. The devil is in the details; the debate should be, too.

author affiliation

Weiland, School of Education, University of Michigan, weilandc@umich.edu

supplemental material

- https://behavioralpolicy.org/journal/
- Supplemental Text
- Additional References

References

1. Public Opinion Strategies & Hart Research. (2014). *Key findings from a national survey of 800 registered voters conducted May 28–June 1, 2014*. Retrieved from Grow America Stronger website: http://growamericastronger.org/wp-content/uploads/2014/06/FINAL_FFYF-National-Results_071414.pdf?24573d

2. Pérez-Peña, R., & Rich, M. (2014, February 3). Preschool push moving ahead in many states. *The New York Times.* Retrieved from http://www.nytimes.com

3. Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2016). *Cradle to kindergarten: A new plan to combat inequality.* Manuscript in preparation.

4. Desilver, D. (2014). For most workers, real wages have barely budged for decades. Retrieved from http://www.pewresearch.org/fact-tank/2014/10/09/for-most-workers-real-wages-have-barely-budged-for-decades/

5. Child Care Aware of America. (2014). *Parents and the high cost of child care: 2014 report*. Arlington, VA: Author.

6. Noss, A. (2014). *Household income: 2013* (American Community Survey Brief 13-02). Washington, DC: United States Census Bureau.

7. Currie, J. M. (1997). Choosing among alternative programs for poor children. *The Future of Children, 7*(2), 113–131.

8. Whitehurst, G. J. (2014). *Whitehurst testimony on early childhood education to the House Committee on Education and the Workforce*. Retrieved from Brookings Institution website: http://www.brookings.edu/research/papers/2014/02/05-congressional-testimony-preschool-whitehurst

9. Yoshikawa, H. (2014). *Testimony to the Senate HELP Committee full-committee hearing on supporting children and families through investments in high-quality early education.* Retrieved from http://www.help.senate.gov/imo/media/doc/Yoshikawa.pdf

10. Barnett, W. S., Friedman-Krauss, A. H., Gomez, R. E., Horowitz, M., Weisenfeld, G. G., & Squires, J. H. (2016). *The State of Preschool 2015: State Preschool Yearbook.* New Brunswick, NJ: National Institute for Early Education Research.

11. U.S. Department of Health and Human Services, Administration for Children and Families, Early Childhood Learning and Knowledge Center, Office of Head Start. (2015). *Head Start program facts: Fiscal year 2014*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/factsheets/2014-hs-program-factsheet.html

12. Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2015). *Cradle to kindergarten: A new plan to combat inequality.* Manuscript in preparation.

13. Clifford, R. M., Barbarin, O., Chang, F., Early, D. M., Bryant, D., Howes, C., . . . Pianta, R. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science, 9,* 126–143.

14. Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children, 5*(3), 25–50.

15. Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15,* 213–238.

16. Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2006). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND.

17. Yoshikawa, H., Weiland, C., & Brooks-Gunn, J. (In press). When does preschool matter? *Future of Children.* 18. Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27,* 109–132.

19. Schindler, H. S., Kholoptseva, J., Oh, S. S., Yoshikawa, H., Duncan, G. J., Magnuson, K.A., & Shonkoff, J. P. (2015). Maximizing the potential of early childhood education to prevent externalizing behavior problems: A meta-analysis. *Journal of School Psychology, 53,* 243–263.

20. McCoy, D. C., Yoshikawa, H., Ziol-Guest, K., Duncan, G. J., Schindler, H., Magnuson, K., . . . Shonkoff, J. P.(2015). *Long-term impacts of early childhood education programs on high school graduation, special education, and grade retention: A meta-analysis*. Manuscript submitted for publication.

21. Li, W., Duncan, G. J., Magnuson, K., Schindler, H., Yoshikawa, H., Leak, J., & Shonkoff, J. P. (2016). *Is timing everything? How early childhood education program impacts vary by starting age, program duration and time since the end of the program*. Manuscript submitted for publication.

22. Zhai, F., Raver, C. C., & Jones, S. (2012). Academic performance of subsequent schools and impacts of early interventions: Evidence from a randomized controlled trial in Head Start settings. *Children and Youth Services Review, 34,* 946–954.

23. Swain, W. A., Springer, M. G., & Hofer, K. G. (2015). Early grade teacher effectiveness and pre-K effect persistence. *AERA Open, 1*(4), Article 2332858415612751.

24. Jenkins, J., Watts, T. W., Magnuson, K., Clements, D., Sarama, J., Wolfe, C., & Spitler, M. (2015). *Preventing preschool fadeout through instructional intervention in kindergarten and first grade* [Working paper]. Retrieved from http://inid.gse.uci.edu/files/2011/03/Jenkinsetal_Fadeout_SREE.pdf

25. Bailey, D., Duncan, G., Odgers, C., & Yu, W. (2015). *Persistence and fadeout in the impacts of child and adolescent interventions* (Life Course Centre Working Paper 2015-27). Retrieved from http://www.lifecoursecentre.org.au/wp-content/uploads/2015/11/2015-27-LCC-Working-Paper-Bailey-et-al.1.pdf

26. Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6,* 142–157.

27. Garces, E., Currie, J., & Thomas, D. (2002). Longer term effects of Head Start. *American Economic Review, 92,* 999–1012.

28. Schweinhart, L. J., Barnett, W. S., & Belfield, C. R. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40.* Ypsilanti, MI: High/Scope Press.

29. Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1,* 111–134.

30. Gibbs, C., Ludwig, J., & Miller, D. L. (2011). *Does Head Start do any lasting good?* (NBER Working Paper No. 17452). Cambridge, MA: National Bureau of Economic Research.

31. Barnett, W. S., & Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review, 26,* 113–125.

32. Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics, 94,* 114–128. doi:10.1016/j.jpubeco.2009.11.001

33. Temple, J. A., & Reynolds, A. J. (2007). Benefits and costs of investments in preschool education: Evidence from the Child–Parent Centers and related programs. *Economics of Education Review, 26,* 126–144.

34. Currie, J. (2001). Early childhood education programs. *Journal of Economic Perspectives, 15,* 213–238.

35. Ludwig, J., & Phillips, D. A. (2008). Long-term effects of Head Start on low-income children. In S. G. Kaler & O. M. Rennert (Eds.), *Annals of the New York Academy of Sciences: Vol. 1136. Reducing the impact of poverty on health and human services: Scientific approaches* (pp. 257–268). New York, NY: New York Academy of Sciences.

36. Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible

explanations. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–116). New York, NY: Russell Sage Foundation.

37. Bassok, D., Finch, J. E., Lee, R., Reardon, S.F., & Waldfogel, J. (2015). *Socioeconomic gaps in early childhood experiences, 1998 to 2010* (EdPolicy Works Working Paper Series No. 38). Retrieved from University of Virginia, Curry School of Education, website: http://curry.virginia.edu/uploads/resourceLibrary/38_Gaps_In_Early_Experience.pdf

38. Bartik, T. J., Gormley, W., & Adelstein, S. (2012). Earnings benefits of Tulsa's pre-K program for different income groups. *Economics of Education Review, 31,* 1143–1161.

39. Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* New York, NY: Foundation for Child Development, Society for Research in Child Development.

40. Heckman, J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). Supplement to "Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program": Web appendices. Retrieved from National Bureau of Economic Research website: http://www.nber.org/data-appendix/w16238/general_apx_2010-07-22a_cji.pdf

41. Puma, M., Bell, S., Cook, R., & Heid, C. (2010). *Head Start impact study: Final report.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

42. Lipsey, M. W., Farran, D. C., & Hofer, K. (2015). *A randomized control trial of the effects of a statewide voluntary prekindergarten program on children's skills and behaviors through third grade.* Retrieved from Peabody Research Institute website: http://peabody.vanderbilt.edu/research/pri/VPKthrough3rd_final_withcover.pdf

43. Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41,* 872–884. doi:10.1037/0012-1649.41.6.872

44. Gormley, W. T., Phillips, D. A., Newmark, K., Welti, K., & Adelstein, S. (2011). Social-emotional effects of early childhood education programs in Tulsa. *Child Development, 82,* 2095–2109.

45. Hill, C. J., Gormley, W. T., & Adelstein, S. (2015). Do the short-term effects of a high-quality preschool program persist? *Early Childhood Research Quarterly, 32,* 60–79.

46. Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009.* Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

47. U.S. Department of Health and Human Services, Administration for Children and Families, Early Childhood Learning and Knowledge Center, Office of Head Start. (2014). *A national overview of grantee CLASS® scores in 2014*. Retrieved from http://eclkc.ohs.acf.hhs.gov/hslc/data/class-reports/class-data-2014.html

48. Phillips, D. A., Gormley, W. T., & Lowenstein, A. E. (2009). Inside the pre-kindergarten door: Classroom climate and instructional time allocation in Tulsa's pre-K programs. *Early Childhood Research Quarterly, 24,* 213–228.

49. Farran, D. C., Hofer, K., Lipsey, M., & Bilbrey, C. (2014, March). *Variations in the quality of TN-VPK classrooms*. Presented at the Society for Research on Educational Effectiveness, Washington, DC.

50. Yoshikawa, H., Leyva, D., Snow, C. E., Treviño, E., Barata, M., Weiland, C., . . . Arbour, M. C. (2015). Experimental impacts of a teacher professional development program in Chile on preschool classroom quality and child outcomes. *Developmental Psychology, 51,* 309–322.

51. Weiland, C., & Yoshikawa, H. (2013). Impacts of a pre-kindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84,* 2112–2130.

52. Bierman, K. L., Domitrovich, C. E., Nix, R. L., Gest, S. D., Welsh, J. A., Greenberg, M. T., . . . Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development, 79,* 1802–1817.

53. Reardon, S. F., & Portilla, X. A. (2015). *Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry* (CEPA Working Paper No. 15-02). Available from Stanford Center for Education Policy Analysis: http://cepa.stanford.edu/wp15-02

54. Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., . . . Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.

55. Sarama, J., Lange, A. A., Clements, D. H., &Wolfe, C. B. (2012). The impacts of an early mathematics curriculum on oral language and literacy. *Early Childhood Research Quarterly, 27,* 489–502.

56. Bierman, K. L., Nix, R. L., Greenberg, M. T., Blair, C., & Domitrovich, C. E. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology, 20,* 821–843.

57. Hulsey, L. K., Aikens, N., Kopack, A., West, J., Moiduddin, E., & Tarullo, L. (2011). *Head Start children, families, and programs: Present and past data from FACES* (OPRE Report 2011-33a). Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

58. Henry, G. T., Ponder, B. D., Rickman, D. K., Mashburn, A. J., Henderson, L. W., & Gordon, C. S. (2004). *An evaluation of the implantation of Georgia's pre-K program: Report of the findings from the Georgia early childhood study (2002–2003).* Retrieved August 25, 2016, from https://www.researchgate.net/profile/Gary_Henry/publication/237436255_AN_EVALUATION_OF_THE_IMPLEMENTATION_OF_GEORGIA'S_PRE-K_PROGRAM_REPORT_OF_THE_FINDINGS_FROM_THE_GEORGIA_EARLY_CHILDHOOD_STUDY_(2002-03)/links/00b495329bd6b45455000000.pdf

59. Duncan, G. J., Jenkins, J. M., Auger, A., Burchinal, M., Domina, T., & Bitler, M. (2015). *Boosting school readiness with preschool curricula.* Retrieved from Irvine Network on Interventions in Development website: http://inid.gse.uci.edu/files/2011/03/Duncanetal_PreschoolCurricula_March-2015.pdf

60. U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). *What Works Clearinghouse™ intervention report: The Creative Curriculum® for Preschool, fourth edition.* Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_creativecurriculum_030513.pdf

61. National Institute for Early Education Research. (2009). *State-approved comprehensive curricula for use in pre-K.* Retrieved September 29, 2015, from http://nieer.org/resources/yearbook/20100621_yearbook_curriculum_data.html

62. Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–31). Baltimore, MD: Brookes.

63. Weiland, C. (In press). Impacts of the Boston Prekindergarten Program on the school readiness of young children with special needs. *Developmental Psychology.*

64. Weiland, C., & Yoshikawa, H. (2014). Does higher peer socio-economic status predict children's language and executive function skills gains in prekindergarten? *Journal of Applied Developmental Psychology, 35,* 422–432.

65. Marshall, N., Roberts, J., & Mills, L. (2006). *Boston Public Schools K1 and K2 programs needs assessment.* Wellesley, MA: Wellesley Centers for Women.

66. Duncan, G., & Murnane, R. (2014). *Restoring opportunity: The crisis of inequality and the challenge for American education.* Cambridge, MA, and New York, NY: Harvard Education Press and Russell Sage Foundation.

67. Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly, 28*(2), 199-209.

A 10-year strategy of increased coordination & comprehensive investments in early child development

Ajay Chaudry & Jane Waldfogel

Summary.   The United States has a fragmented and inadequate system of early childhood care and education. Most children do not receive high-quality opportunities; only those whose families have the means to provide them receive their benefits. Market-based and privately financed services operate separately from publicly funded programs and serve different populations with different levels of quality. Often, there is little coordination between different programs that serve the same child over the course of her or his early years. This lack of coordination contributes to growing inequities in later educational and adult outcomes. We propose a 10-year strategy for a coordinated set of reforms to significantly improve and integrate the major public and private early childhood programs into a coherent whole. The goal is to better meet children's needs, with a special focus on leveling the development and learning gaps that exist before kindergarten. The strategy consists of paid parental leave, child-care assistance for children with working parents, universal early education that starts when children are 3 years old, and a re-envisioned role for Head Start to reach the most disadvantaged children with intensive services from birth.

The United States invests far less in children between birth and age 5 years than in older children and lags far behind what other countries with well-developed economies spend on early childhood programs.[1,2] For example, in 2012, 54% of 3- and 4-year-olds in the United States were enrolled in any type of preschool education, compared with 94% of 3- and 4- year-olds in Germany, 96% in the United Kingdom, and 100% in France. While U.S. families paid for the early education of more than half of the children enrolled in programs, more than 80% of children in Germany, the United Kingdom, and France were served through public funds. In addition, although the United States, on average, spends approximately 10% more on elementary and secondary education (kindergarten through 12th

grade; K–12) as these countries, it spends less than half of what they do on preschool education.[1]

In contrast to these peer countries that provide universal programs, the United States provides deeply fragmented early childhood education (ECE) experiences for its children. Among high-income families, 76% of 3- and 4-year-olds attend preschool; of these, 75% are in private programs. At the same time, among children in low- and middle-income families, only about half (49% and 54%, respectively) attend preschool, with the majority participating in publicly funded programs (80% of children from low-income families and 54% of those from middle-income families). For those who enroll in public programs, there is a range of different service systems (such as the federal Head Start program, state-funded prekindergarten programs in some states, and federal and state-funded child-care subsidies), but each system has its own set of eligibility rules and quality standards, and none are funded to the degree needed to fully meet the needs of the eligible population they are intended to serve. For children younger than 3 years, the gaps are even wider. As a result, on the first day of kindergarten, many American children are not ready for school.

The fragmented and inadequate investment in early childhood programs has continued despite an abundance of research across the biological, economic, and behavioral sciences showing that children's cognitive, social, and emotional skills develop the most rapidly during their early years and that this is a time when investments are particularly effective.[4–6] Failing to invest adequately in young children's skill development hurts the ability of the United States to develop the highly skilled labor force that is crucial for competing in the global economy.

The wide disparities in young children's cognitive and social-emotional skills by parents' educational and income levels increase later educational and economic inequalities. Large disparities in reading and math skills between children from low or medium socioeconomic status families and children from higher socioeconomic status families are already present at the start of kindergarten. These disparities have been widening in recent decades[8,10] and have lifelong consequences for children's outcomes.[7,11,12] Studies that track children longitudinally find that reading and math scores in kindergarten and the early primary grades strongly predict later outcomes such as high school graduation,

college attendance, college completion, adult earnings, health, and criminal behavior.[11,13,14] Studies also show that preschool programs can have lasting effects. For example, David Deming found that, compared with control participants, Head Start participants are over 8 percentage points more likely to graduate from high school, 6 percentage points more likely to attend college, 2 percentage points less likely to become teen parents, and 7 percentage points less likely to be in poor health in young adulthood.[13]

In this article, we make the case that it is urgent to address the wide and growing disparities in young children's cognitive and social-emotional skills by creating a comprehensive set of investments in early childhood. Many proposals for ECE focus on reforms to individual service systems (for example, prekindergarten or home visiting), a narrow population of children (for example, 4-year-olds or children of working parents), or a particularly significant problem in an existing service (for example, low-quality child care). We are concerned that, in isolation, these efforts might prove insufficient or could fragment services even more. We also believe that it is time for the United States, like the peer countries mentioned above, to move to a universal system of early care and education. Providing universal care and education in early childhood benefits all children, not just the disadvantaged. And, if the goal is to provide high-quality services to low-income children and early experiences on a par with those received by their higher-income peers, that goal is more likely to be achieved if middle- and high-income children are able to participate in universal programs and benefits.

A strong evidence base supports the effectiveness of early childhood interventions, including many studies using randomized controlled trials or other rigorous methods, enough for the country to move forward with needed reforms.[15,16] At the same time, further research is needed to test rigorous models that can be developed at a significant and replicable scale.

In the next section, we discuss the principles that guide our approach. Following that, we lay out the four components of our 10-year strategy: *paid parental leave* to ensure quality caregiving from birth, *guaranteed child-care assistance* for families with working parents; *universal early education starting when children are 3 years old*; and a *re-envisioned role for Head Start* that focuses on the most vulnerable children in the most disadvantaged communities.

Principles to Guide Early Childhood Investments

Our approach is grounded in four principles based on what we see as the key challenges across the range of early childhood services and the core societal values that are relevant to this area.

1. *Access.* Ensure that all American children have access to the opportunities they need to reach their developmental potential and embark on a secure educational pathway. This means supporting evidence-based early childhood services that increase the social, emotional, and intellectual development of children from birth to age 5 years.

2. *Quality.* Public investments should provide high-quality care and education opportunities. Research indicates that higher quality care and education provide lasting benefits for children's development and outcomes[15,17,18] but that much of the early care and education children receive currently is low in quality.[19,20] High quality in early care and education is generally defined in terms of the program characteristics associated with positive effects on children's cognitive, social, and emotional development. These include both what are considered structural characteristics (such as the training and skill level of the teacher or caregiver and appropriate teacher-to-child ratios at given ages for children) as well as process quality characteristics (such as the amount and type of direct interactions children experience with their caregivers).

3. *Parental support.* Public investments must support parents, who have the primary responsibility for nurturing their children's early development. Parents have the greatest influence on children, both directly as a result of their parenting styles and indirectly through the early learning situations in which they place their children.[21] This means that in addition to offering support for parenting skills, policymakers can help parents make good choices by improving the availability, quality, and affordability of developmental opportunities.

4. *Shared private and public responsibility.* All parents should be expected to contribute to the extent their resources permit, because they have the greatest stake in their children's future and are best positioned to make decisions regarding the most appropriate investments. At the same time, what parents can provide is constrained by their resources relative to the costs of high-quality care and education. In some instances in which investments support a service that responds to broadly shared needs (for example, for public education or parental leave to care for newborn children), public financing and universal provision of services may offer the best approach. In other instances, it may be more efficient to expect market-based services to meet a range of family preferences and needs, supplemented by targeted public support to ensure access or assure quality.

A Four-Pronged, 10-Year Strategy to Expand and Coordinate Early Childhood Investments

On the basis of the evidence about what children need and effective interventions to meet these needs, we propose a coordinated strategy for investments in early childhood. Our approach builds on extensive research on children's early development, the effectiveness of program interventions, and the role of public policies in enhancing the development of the nation's children, including its most vulnerable children.[15,22,23]

Parental Leave: Supporting Children and Parents from Birth

We propose that the United States institute paid parental leave as a national policy. The goal should be to provide a minimum of 12 to 16 weeks of paid leave to all new parents after the birth or adoption of a child.

The days and weeks following birth are a critical time when a newborn needs consistent, sensitive, and responsive care. However, many parents in the United States are unable to take time off from work to care for their children without risking the loss of their job or their family income. The United States is the only developed country that does not provide public financing for paid leave to parents with a newborn baby. As a result, three in 10 first-time mothers in the United States return to the workplace within 2 months of their baby's birth, and those with the least resources—those who are less educated, single, or younger—are the most likely to return to work early.[24]

Extensive research elucidates the benefits of more generous parental leave policies. Several studies have shown that a quick return to work after childbirth is associated with early cessation of breastfeeding, more maternal depression, and poorer child health outcomes.[22,25] The research to date on the potential

benefits of such policies has largely been conducted in other countries or in the few states that have implemented paid leave programs (California, New Jersey, and Rhode Island). These studies suggest several benefits for children and families, including longer durations of breastfeeding,[26] higher rates of vaccination and well-baby doctor visits,[22,27] lower likelihood of mothers becoming depressed,[28] and improved educational outcomes for children.[29]

Funding the reform

We propose providing paid parental leave as a form of federal social insurance through the Social Security system or a similar mechanism. There are several options for funding this insurance. One would be to marginally increase payroll taxes for employers and employees. Another approach would be to fund parental leave through general revenues as part of a tax reform effort, perhaps by limiting the value of itemized deductions for higher income taxpayers. (In drawing up the federal policy, Congress can learn from the three states that currently have paid leave policies, additional cities and states that are innovating in this area, and employers that have implemented parental leave programs.)

Assuring Families with Young Children Access to High-Quality Child Care

Given that most children's parents work, nonparental care is a common experience of young children in the United States today. Many parents struggle to find and afford high-quality care, especially for infants and toddlers. Yet studies using standardized and widely applied measures of quality[3,19,30] find that most of the care received by American children during this period of rapid brain development[6,31] is of low or mediocre quality.

Children from low- and middle-income families, on average, experience lower quality care, whereas higher income families primarily use higher quality early care and education arrangements that are unavailable or unaffordable for less affluent families.[3,32] As a result, even though they spend a substantial proportion of their earnings on early care and education, low- and middle-income families are generally unable to purchase the highest quality services.

Although the United States does have programs that assist with child-care costs and help parents start or

continue to work,[33] they are limited in their reach and their level of support. The primary support for low-income families is the federal- and state-funded Child Care and Development Fund (CCDF), which combines federal block grants and state matching funds. Because of limited public funding, CCDF served only 15% of eligible low-income families in 2012.[34] The subsidies are so tightly rationed that many low-income working families do not get any assistance, whereas others get low-quality care, often for very short durations.[35,36] Two additional federal tax programs—the Child and Dependent Care Tax Credit (CDCTC) and Dependent Care Assistance Plan—reach more people, but because only households with income tax liability are eligible for the benefits, poor and low-income households usually do not qualify for these programs. Even those who receive support get a modest subsidy relative to the costs of child care.[37]

Funding the reform

Both the subsidy and the tax programs need to be updated and significantly expanded to better support low- and moderate-income families' access to high-quality, affordable child care. First, the federal government and the states should guarantee subsidies for licensed child care to low- and moderate-income families (that is, families with incomes of up to 250% of the federal poverty line, which was $60,625 for a family of four in 2015) in which all parents residing with children under the age of 5 years are working. Second, the federal CDCTC should be expanded to provide support for a wider range of types of care and family income levels than the subsidy program covers. That is, it should include families whose incomes are above 250% of the federal poverty line for whom the cost of high-quality child care would still represent an unaffordable expense. Families would be expected to pay a proportion of their earnings for the cost of their children's care, and this proportion would increase progressively with income.

Begin Universal ECE when Children Are 3 Years Old

Mounting evidence in neuroscience, developmental psychology, and economics has shown how ECE can enhance children's skill development, school readiness, and longer term educational attainment and employment trajectories, with children receiving higher quality education and care demonstrating greater gains.[6,13,17,38]

Understanding the importance of early learning, parents have been enrolling their children in school- or center-based ECE (more commonly referred to as *preschool*) at increasing rates and at earlier ages over the last 25 years. By 2013, 4.7 million 3- and 4-year-olds (that is, more than half of the children in this age group) attended preschool. However, as noted earlier, both access to preschool and its quality vary by family income.[3] Publicly funded preschool increases access to ECE and narrows income-based disparities in access, but sizeable gaps still persist for children in both low- and middle-income homes. The majority of states spend a fraction of what is expended on K–12 education for preschool-age education, although the cost of ECE should generally be higher given the lower adult-to-child ratios in preschool.

Recent rigorous studies of large-scale public preschool programs in Boston and Tulsa—which have been identified as relatively high-quality programs— show substantial gains for a range of school readiness outcomes.[39,40] These studies also indicate that ECE is beneficial for children in families of all income levels, with the greatest benefits accruing to children from more disadvantaged economic backgrounds, thus identifying ECE as a key opportunity to reduce the current sizeable gaps in school readiness. However, a recent evaluation of Tennessee's prekindergarten program found much smaller initial benefits, and these were no longer evident by the end of first grade. The findings may reflect the lower quality of the services in the Tennessee program, given that 85% of classrooms in a representative sample were found not to meet a standard of "good" overall quality.[41] These results suggest that investing significantly to raise the quality of ECE programs is as critical as expanding access, particularly for children from low-income families, who stand to gain the most in reading and math from higher quality programs. Although delivering quality at scale remains a challenge, results from Boston indicate that the adoption of evidence-based, developmentally focused curricula together with classroom-based coaching may be one promising route to improving program quality.[42,43]

As most higher income families now enroll their children in preschool by the time their children are 3 years old, the disparities in school readiness between children from high-income families and children from low- and middle-income families have widened. Given the documented benefits of high-quality preschool for all children,[15,44] we believe that high-quality, free, universal, public ECE in the United States should begin for children at the age of 3 years, to ensure all children have a common educational starting point.

Funding the Reform

Consistent with state and local government responsibility for primary and secondary education, we envision state and local governments taking the primary responsibility for funding, developing, and implementing universal early education. Many states are already investing in ECE because they believe that it will be cost-effective and improve educational outcomes in the long run. But it is also true that many states are struggling to find the funding to expand ECE given other pressing needs. Therefore, the federal government may need to support states through matching funds over the next 10 years to accelerate the building of high-quality, well-sequenced ECE programs that are integrated with elementary education. The federal government might also provide resources for research and development.

Remaking Head Start

The above three components will meet the needs of many American children in early childhood. But what about the most disadvantaged? Studies over several decades have documented how childhood poverty and low levels of family income and resources lead to worse child outcomes, particularly in cognitive and educational areas.[45,46] Furthermore, early experiences of poverty;[47] persistent, chronic poverty;[48] and higher concentrations of poverty in the community[49] have been found to lead to even worse child outcomes than the experience of poverty alone.

Research indicates that intensive high-quality programs targeted to this population can address some of these disadvantages and improve child health and development. For example, the Abecedarian Project, a model program that served children in one highly impoverished community in North Carolina in the 1970s, showed very large positive effects for participants both in their school careers and through young adulthood on a wide range of important outcomes. Compared with a control group, Abecedarian participants were more likely to earn a 4-year college degree (23% versus 6%), more likely to be employed full time

at age 30 years (75% versus 53%), and less likely to become teen parents (26% versus 45%). They also showed significantly better health measurements (11% of participants had high blood pressure in their mid-30s compared with 44% for the control group; 11% of participants were severely obese in their mid-30s compared with 38% of the control group).[50–52] Another program, the Infant Health and Development Program (IHDP), built on the Abecedarian Project's successful model at a larger scale by serving low-birth-weight newborns across eight urban study sites in the 1990s, offering frequent home visitation in the first year and full-day, high-quality early education beginning when infants were 1 year old and continuing until they turned 3 years old. IHDP significantly raised the cognitive skills of children who were from low-income families, such that income-based gaps in school readiness at age 5 years were substantially reduced.[53]

The federal Early Head Start program—which offers home visiting and/or center-based care for children up to 3 years old—is the closest program to IHDP that has been implemented on a national scale, although its services are much less intensive and less consistently high quality.[54] An evaluation of Early Head Start in its initial years of implementation showed modest benefits in terms of improvements in children's cognitive and social-emotional development.[55]

There are also several proven parenting programs.[56,57] Although home-visiting programs have had a mixed track record, some have been found to be very effective in improving parenting and child outcomes in disadvantaged families.[58] For example, substantial evidence supports the benefits of the Nurse-Family Partnership program, which provides a series of home visits by trained nurses to disadvantaged first-time mothers in pregnancy and the first 2 years after they give birth.[59] Several other model programs have also been found to provide benefits in a comprehensive federal review of programs subject to rigorous evaluation.[60]

On the basis of these findings, we propose a narrowly targeted, intensive, and comprehensive child development initiative for children who begin life in concentrated poverty or face particular adversities that need much more focused early attention and services. We believe that the nation's current Head Start and Early Head Start programs, together with the federal evidence–based home visiting programs,

can be reimagined and reintegrated to target the most disadvantaged children and families. Among the goals of this initiative is to help parents increase their ability to support the development of their children. Another goal is to provide early screening (in collaboration with Medicaid and pediatric services) to identify children with multiple and serious disadvantages and give them services or arrange services though community-based programs. Placed within the most disadvantaged geographical areas, Head Start centers would serve as community-based comprehensive service hubs that coordinate early learning with parenting supports, home-visiting services, specialized early intervention services, and medical care.

Funding the reform
Shifting Head Start to serve younger children, in conjunction with the expansion of universal preschool in the states, would require additional federal resources for Head Start even if the total number of children served through Head Start were held constant, because programs for younger children require greater funding per child.

Conclusion

The large and growing inequalities in the earliest years of a child's life, if not addressed, will add to the broader educational and economic inequalities that are a defining problem for this century. A huge gulf exists between the current set of fragmented and limited investments in early child development in the United States and what is needed for all children to have the opportunity to succeed.

The coordinated approach we propose would significantly change and integrate the current major public and private early childhood programs into a coherent whole to better meet the needs of all children while helping level the early development and learning gaps that are seen before kindergarten. Although better integration of existing services would make a substantial difference, to truly even come close to meeting the needs of all young children would require considerable additional investments, likely a multiple of what is spent now, which would be more in line with what is spent in other developmental stages (such as K–12 and postsecondary education).

We recognize that our proposal is costly. Raising public spending in early childhood to the roughly $12,400 average expenditure per year per child in the United States[61] on education for school-age children would require an additional $223 billion annually, or 1.3% of the gross domestic product. Although our proposals would not amount to this much when fully implemented in 10 years, we believe that the United States cannot simply tinker at the edges. Because funds will be limited for the foreseeable future, we recommend starting with what is most needed or most ready for investment. That would include setting up publicly funded paid parental leave; implementing universal preschool, starting with 4-year-olds; and beginning to redirect Head Start funds to serve the most disadvantaged children starting at birth.

We note that there is also a need for investments to help reduce poverty and financial insecurity among families with children. We have not discussed important measures such as expanded child tax credits or a child allowance here, but we do acknowledge the need for such instruments.[60]

It is our belief that together, such investments will make a meaningful difference in improving the life prospects of young children in America. If the United States invests in providing opportunities to overcome disadvantaged starting points, more children will be able to live up to their potential.

author affiliation

Chaudry, Robert F. Wagner Graduate School of Public Service & Institute for Human Development and Social Change, New York University; Waldfogel, School of Social Work and Columbia Population Research Center, Columbia University. Corresponding author's e-mail: ac1154@nyu.edu

supplemental material

- https://behavioralpolicy.org/journal/
- Supplemental Text

References

1. Organization for Economic Co-operation and Development. (2015). *Education at a glance 2015: OECD indicators.* LOCATION: Author.
2. Organization for Economic Co-operation and Development. (2015). *Enrollment in child care and preschool: OECD indicators.* LOCATION: Author.
3. Chaudry, A., Morrissey, T., Weiland, C., & Yoshikawa, H. (2016). *A strategic vision for addressing inequality in early childhood.* Manuscript in preparation.
4. Cunha, F., & Heckman, J. J. (2007). The technology of skill formation. *American Economic Review, 97*(2), 31–47.
5. Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development, 81,* 357–367.
6. Knudsen, E. I., Heckman, J., Cameron, J., & Shonkoff, J. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, USA, 103,* 10155–10162.
7. Duncan, G. J., & Magnuson, K. (2011). The nature and impact of early achievement skills, attention skills, and behavior problems. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 47–70). New York, NY: Russell Sage Foundation.
8. Reardon, S. F. (2011). The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In G. Duncan & R. Murnane (Eds.), *Whither opportunity: Rising inequality, schools, and children's life chances* (pp. 91–116). New York, NY: Russell Sage Foundation.
9. Bradbury, B., Corak, M., Waldfogel, J., & Washbrook, E. (2015). *Too many children left behind: The U.S. achievement gap in comparative perspective.* New York, NY: Russell Sage Foundation.
10. Reardon, S. F., & Portilla, X. A. (2015). *Recent trends in socioeconomic and racial school readiness gaps at kindergarten entry* (CEPA Working Paper No. 15-02). Available from Stanford Center for Education Policy Analysis website: https://cepa.stanford.edu/content/recent-trends-income-racial-and-ethnic-school-readiness-gaps-kindergarten-entry
11. Chetty, R., Friedman, J., Hilger, N., Saez, E., Schanzenbach, D., & Yagan, D. (2010). *How does your kindergarten classroom affect your earnings? Evidence from Project STAR* (NBER Working Paper No. 16381). Available from National Bureau of Economic Research website: http://www.nber.org/papers/w16381
12. Heckman, J. J., Moon, S., Pinto, R., Savelyev, P., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics, 94,* 114–128. doi:10.1016/j.jpubeco.2009.11.001
13. Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics, 1,* 111–134.
14. Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A., Klebanov, P., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43,* 1428–1446.
15. Yoshikawa, H., Weiland, C., Brooks-Gunn, J., Burchinal, M. R., Espinosa, L. M., Gormley, W., & Zaslow, M. J. (2013). *Investing in our future: The evidence base on preschool education.* New York, NY: Foundation for Child Development.
16. Ruhm, C., & Waldfogel, J. (2012). Long-term effects of early childhood care and education. *Nordic Economic Policy Review, 1,* 23–52.

17. McCoy, D. C., Yoshikawa, H., Ziol-Guest, K., Duncan, G. J., Schindler, H., Magnuson, K., . . . Shonkoff, J. P. (2015). *Long-term impacts of early childhood education programs on high school graduation, special education, and grade retention: A meta-analysis.* Manuscript submitted for publication.

18. Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–31). Baltimore, MD: Brookes.

19. Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O., Bryant, D., & Howes, C. (2008). Measures of classroom quality in prekindergarten and children's development of academic, language, and social skills. *Child Development, 79,* 732–749.

20. Moiduddin, E., Aikens, N., Tarullo, L., West, J., & Xue, Y. (2012). *Child outcomes and classroom quality in FACES 2009.* Washington, DC: U.S. Department of Health and Human Services.

21. Waldfogel, J., & Washbrook, E. (2011). Income-related gaps in school readiness in the United States and the United Kingdom. In T. Smeeding, R. Erickson, & M. Jäntti (Eds.), *Persistence, privilege, and parenting: The comparative study of intergenerational mobility* (pp. 175–208). New York, NY: Russell Sage Foundation.

22. Berger, L. M., Hill, J., & Waldfogel, J. (2005, February). Maternity leave, early maternal employment, and child health and development in the U.S. *The Economic Journal, 115*(501), F29–F47.

23. Duncan, G. J., & Magnuson, K. (2013). Investing in preschool programs. *Journal of Economic Perspectives, 27,* 109–132.

24. Laughlin, L. (2011). *Maternity leave and employment patterns of first-time mothers: 1961–2008* (Current Population Report P70-128). Retrieved from U.S. Census Bureau website: http://www.census.gov/prod/2011pubs/p70-128.pdf

25. Guendelman, S., Kosa, J. L., Pearl, M., Graham, S., Goodman, J., & Kharrazi, M. (2009). Juggling work and breastfeeding: Effects of maternity leave and occupational characteristics. *Pediatrics, 123,* e38–e46.

26. Appelbaum, E., & Milkman, R. (2011). *Leaves that pay: Employer and worker experiences with paid family leave in California.* Washington, DC: Center for Economic and Policy Research.

27. Daku, M., Raub, A., & Heymann, J. (2012). Maternal leave policies and vaccination coverage: A global analysis. *Social Science & Medicine, 74,* 120–124.

28. Chatterji, P., & Markowitz, S. (2012). Family leave after childbirth and the mental health of new mothers. *Journal of Mental Health Policy and Economics, 15,* 61–76.

29. Corneiro, P., Løken, K. V., & Salvanes, K. G. (2011). *A flying start? Maternity leave benefits and long run outcomes of children* (IZA Discussion Paper No. 5793). Retrieved from Institute for the Study of Labor website: http://ftp.iza.org/dp5793.pdf

30. Weiland, C., Ulvestad, K., Sachs, J., & Yoshikawa, H. (2013). Associations between classroom quality and children's vocabulary and executive function skills in an urban public prekindergarten program. *Early Childhood Research Quarterly, 28,* 199–209.

31. Shonkoff, J. P., & Phillips, D. A. (Eds.). (2000). *From neurons to neighborhoods: The science of early childhood development.* Washington, DC: National Academy Press.

32. Dowsett, C. J., Huston, A. C., Imes, A. E., & Gennetian, L. (2008). Structural and process features in three types of child care for children from high and low income families. *Early Childhood Research Quarterly, 23,* 69–93.

33. Crawford, A. (2006). The impact of child care subsidies on single mothers' work effort. *Review of Policy Research, 23,* 699–711.

34. Chien, N. (2015). *Estimates of child care eligibility and receipt for fiscal year 2012* [Issue brief]. Retrieved from U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation website: https://aspe.hhs.gov/sites/default/files/pdf/153591/ChildEligibility.pdf

35. Henly, J., Sandstrom, H., Claessens, A., Pilarz, A., Gelatt, J., Kim, J., & Healy, O. (2015). *Determinants of subsidy stability and child care continuity: Final report for the Illinois–New York Child Care Research Partnership.* Retrieved from Urban Institute website: http://www.urban.org/research/publication/determinants-subsidy-stability-andchild-care-continuity

36. Herbst, C., & Tekin, E. (2010). Child care subsidies and child development. *Economics of Education Review, 29,* 618–638.

37. Tax Policy Center. (XXXX). Quick facts: Child and Dependent Care Tax Credit (CDCTC). Retrieved from http://www.taxpolicycenter.org/press/quickfacts_cdctc.cfm

38. Bartik, T. J., Gormley, W., & Adelstein, S. (2012). Earnings benefits of Tulsa's pre-K program for different income groups. *Economics of Education Review, 31,* 1143–1161.

39. Weiland, C., & Yoshikawa, H. (2013). Impacts of a pre-kindergarten program on children's mathematics, language, literacy, executive function, and emotional skills. *Child Development, 84,* 2112–2130.

40. Gormley, W. T., Gayer, T., Phillips, D., & Dawson, B. (2005). The effects of universal pre-K on cognitive development. *Developmental Psychology, 41,* 872–884.

41. Farran, D. C., Hofer, K., Lipsey, M., & Bilbrey, C. (2014, March). *Variations in the quality of TN-VPK classrooms.* Presentation at the Society for Research on Educational Effectiveness, Washington, DC.

42. Clements, D. H., & Sarama, J. (2008). Experimental evaluation of the effects of a research-based preschool mathematics curriculum. *American Educational Research Journal, 45,* 443–494.

43. Morris, P. A., Mattera, S., Cattells, N., Bangser, M., Bierman, K. L., & Raver C. C. (2014). *Impact findings from the Head Start CARES demonstration.* New York, NY: MDRC.

44. Arteaga, I., Humpage, S., Reynolds, A., & Temple, J. (2014). One year of preschool or two: Is it important for adult outcomes? *Economics of Education Review, 40,* 221–237.

45. Dahl, G. B., & Lochner, L. (2012). The impact of family income on child achievement: Evidence from the earned income tax credit. *The American Economic Review, 102,* 1927–1956.

46. Duncan, G. J., Morris P. A., & Rodrigues, C. (2011). Does money really matter? Estimating impacts of family income on young children's achievement with data from random-assignment experiments. *Developmental Psychology, 47,* 1263–1279.

47. Ratcliffe, C., & McKernan, S. (2012). *Child poverty and its lasting consequences.* Washington, DC: Urban Institute.

48. Duncan, G. J., Ziol-Guest, K., & Kalil, A. (2010). Early-childhood poverty and adult attainment, behavior, and health. *Child Development, 81,* 306–325.

49. Wodtke, G. T., Harding, D. J., & Elwert, F. (2011). Neighborhood effects in temporal perspective: The impact of long-term exposure to concentrated disadvantage on high school graduation. *American Sociological Review, 76,* 713–736.

50. Campbell, F., Pungello, E., Burchinal, M., Kainz, K., Pan, Y., Wasik, B., . . . Ramey, C. (2012). Adult outcomes as a function of an early childhood educational program: An Abecedarian Project follow-up. *Developmental Psychology, 48,* 1033–1043.

51. Campbell, F., Ramey, C., Pungello, E., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42–57.

52. Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., Pinto, R., Pungello, E. P., & Pan, Y. (2104, March 28). Early childhood investments substantially boost adult health. *Science, 343,* 1478–1474.

53. Duncan, G. J., & Sojourner, A. J. (2013). Can intensive early childhood intervention programs eliminate income-based cognitive and achievement gaps? *Journal of Human Resources, 48,* 945–968.

54. Love, J., Kisker, E., Ross, C., Raikes, H., Constantine, J., Boller, K., . . . Vogel, C. (2005). The effectiveness of Early Head Start for 3-year-old children and their parents: Lessons for policy and programs. *Developmental Psychology, 41,* 885–901.

55. Love, J., Chazan-Cohen, R., Raikes, H., & Brooks-Gunn, J. (2013). What makes a difference: Early Head Start evaluation findings in a developmental context. *Monographs of the Society for Research in Child Development, 78*(1).

56. Dozier, M., Peloso, E., Lewis, E., Laurenceau, J. P., & Levine, S. (2008). Effects of an attachment-based intervention on the cortisol production of infants and toddlers in foster care. *Development and Psychopathology, 20,* 845–859.

57. Landry, S. H., Smith, K. E., Swank, P. R., & Guttentag, C. (2008). A responsive parenting intervention: The optimal timing across early childhood for impacting maternal behaviors and child outcomes. *Developmental Psychology, 44,* 1335–1353.

58. Avellar, S., Paulsell, D., Sama-Miller, E., Del Grosso, P., Akers, L., & Kleinman, R. (2015). *Home visiting evidence of effectiveness review* [Executive summary]. Washington, DC: U.S. Department of Health and Human Services.

59. Olds, D., Eckenrode, J., Henderson, C., Jr., Kitzman, H., Powers, J., Cole, R., . . . Luckey, D. (1997). Long-term effects of home visitation on maternal life course and child abuse and neglect: 15-year follow-up of a randomized trial. *Journal of the American Medical Association, 278,* 637–643.

60. Garfinkel, I., Harris, D., Waldfogel, J., & Wimer, C. (2016). *Doing more for our children.* NY, New York: Century Foundation.

61. Kena, G., Musu-Gillette, L., Robinson, J., Wang, X., Rathbun, A., Zhang, J., . . . Velez, E. D. (2015). *The condition of education 2015* (NCES 2015-144). Retrieved from the U.S. Department of Education, National Center for Education Statistics website: http://nces.ed.gov/pubs2015/2015144.pdf

Reimagining accountability in K–12 education

Brian P. Gill, Jennifer S. Lerner, & Paul Meosky

Summary. Since the passage of the No Child Left Behind Act (NCLB) in 2002, American policymakers have relied primarily on outcome-based accountability in the form of high-stakes testing to improve public school performance. With NCLB supplanted in 2015 by the Every Student Succeeds Act—which gives states far greater discretion in the design of accountability systems—the time is ripe for policymakers to consider extensive behavioral science literature that shows outcome-based accountability is only one of multiple forms of accountability, each invoking distinct motivational mechanisms. We review rule-based, market-based, and professional accountability alongside outcome-based accountability, using evidence from the laboratory and the field to describe how each can produce favorable or unfavorable effects. We conclude that policymakers should (a) make greater use of professional accountability, which has historically been underutilized in education; (b) use transparency to promote professional accountability; and (c) use multiple, complementary forms of accountability, creating a complete system that encourages and supports the continuous improvement of educational practice.

The Equity Project (TEP) Charter School is a public middle school in the Washington Heights neighborhood of New York City, where, since opening in 2009, it has served a student population that is almost exclusively low income and Hispanic or African American. TEP Charter School's founding principal, Zeke Vanderhoek a former teacher and entrepreneur who had read the growing body of research indicating that teachers are the most important school-based influence on students, decided to design a school that would focus virtually all of its resources on hiring and developing the best possible teachers. Vanderhoek ran the numbers and concluded that the standard per-pupil public funding available to New York City charter schools could be reallocated (for example, by eliminating administrative positions and increasing the size of a typical class from 27 students to 31) in a way that would allow him to pay each of his teachers $125,000 per year, plus bonuses based on school-wide student achievement. The plan worked. In its first 4 years of

operation, TEP Charter School produced substantial positive effects on student achievement: By the time they finished eighth grade, TEP Charter School's initial cohort of fifth graders had test scores showing an advantage equivalent to several months of additional instruction in English and science and more than a year and a half of additional instruction in math, relative to a similar comparison group.[1]

The Equity Project Charter School: A Case Study in Multi-Mode Accountability

Although media coverage of TEP Charter School focused on the eye-popping teacher salaries—which enabled TEP Charter School to hire, among other star educators, a physical education teacher who was previously a trainer for the Los Angeles Lakers—the school's approach involves much more than high salaries and performance-based bonuses. Those who seek to teach at TEP Charter School are rigorously screened in a process that requires applicants to spend a full day teaching TEP Charter School's students. Once hired, TEP Charter School teachers engage in 6 weeks of professional development annually. In addition, throughout the school year, they observe each other in the classroom at least twice a week, providing written feedback to their colleagues. TEP Charter School does not offer tenure, and teachers' contracts are renewed on the basis of their performance, which is evaluated primarily through classroom observation. In short, TEP Charter School staff are accountable for their students' achievement, but they are also accountable for their professional practice, which is observed not only by the principal but also by their colleagues, who are expected to help them improve their practice. In addition, as a charter school, TEP Charter School is implicitly accountable to the parents of its students, because its funding depends on persuading parents to enroll their children.

Accountability as High-Stakes Testing

Since the turn of the millennium, American policymakers have tended to understand accountability in education in narrow terms. The No Child Left Behind Act (NCLB) required reporting of every public school's level of student proficiency in math and reading and imposed sanctions for failure to achieve targets. Schools that fell short of state-determined proficiency goals faced mandatory interventions, which became increasingly aggressive after repeated failures. Chronically failing schools could be restaffed, taken over by the state, or shut down. *Accountability* came to be synonymous with high-stakes testing. The Obama administration went even further in promoting high-stakes testing, pushing states to include student achievement growth measures in evaluating teachers and principals.

High-stakes testing has come under attack on multiple fronts. Teachers unions objected to the use of test scores in teacher evaluations; high-profile cheating scandals ruined the careers of prominent educators; parents increasingly demanded the right to opt out of standardized testing of their children. Meanwhile, education scholars proposed replacing existing accountability regimes.[2–4] Recognizing the increasing unpopularity of the federal mandates for high-stakes testing, in December 2015, Congress passed the Every Student Succeeds Act, which replaces NCLB and gives states far more discretion to design their own accountability policies.

Opportunity to Craft Better Accountability Systems

With new latitude to refine and improve kindergarten through 12th grade (K–12) accountability policies, policymakers should heed the empirical evidence on accountability's effects—and not only the evidence from education itself. Extensive literature in experimental social psychology and behavioral economics identifies different types of accountability mechanisms and the conditions under which accountability improves outcomes.[5] Just as important, it also identifies conditions in which accountability may have no effect or even make matters worse. This behavioral science literature can provide valuable insights for the design of school accountability systems, but it has been largely overlooked in the education policy debate.[6] (The National Academies report cited here examined some of the behavioral literature related to the use of test-based incentives, but it did not examine other forms of accountability.)

> *Accountability comes in many forms, involving different mechanisms and different behavioral responses*

Jennifer S. Lerner of Harvard University and Philip Tetlock, of the University of Pennsylvania, conducted a comprehensive review of the behavioral science literature on accountability;[5] we use that review to provide a framework for our discussion here. Lerner and Tetlock's review made two broad points that motivate our discussion. First, they noted that accountability comes in many forms, involving different mechanisms and different behavioral responses. Outcome-based accountability is only one of these forms. In the context of schooling, this means that high-stakes testing is not the only tool available for improving educational performance. Second, Lerner and Tetlock emphasized that accountability can have positive or negative effects, depending on the accountability type, the decision context, and the nature of the task. This suggests that policymakers need to seriously consider the trade-offs associated with different types of accountability and explore ways to mitigate unintended effects.

Four Types of Accountability

In this article, we explore the behavioral science literature alongside the education literature on accountability. We discuss evidence related to four types of accountability applicable to education policy: **rule-based, market-based, outcome-based,** and **professional** accountability. We also propose how policymakers might use these different types of accountability to promote continuous improvement in schools. Schooling has multiple purposes with multiple constituencies, including parents, students, and the public. Any single form of accountability is unlikely to fully serve these multiple purposes and constituencies.

We reach several conclusions relevant to policy and practice. First, transparency alone can create accountability, even in the absence of explicit rewards and sanctions. Second, multiple forms of accountability can be used concurrently to take advantage of complementary strengths and weaknesses. Third, to promote continuous improvement in schools, a comprehensive accountability system should include mechanisms for the improvement of practice. Collectively, these three points suggest greater use of professional accountability than has historically been the norm in K–12 education, alongside rule-based, market-based, and outcome-based accountability.

Accountability Mechanisms and Types

Four Behavioral Mechanisms for Achieving Accountability

Lerner and Tetlock's review[5] identified four mechanisms that make people feel accountable: (a) the *mere presence of another*—simple awareness that someone else is watching, (b) *identifiability*—the expectation that an action or outcome will be attributable to oneself, (c) *reason giving*—the expectation that one will need to explain or rationalize one's actions, and (d) *evaluation*—the expectation that consequences will depend on an assessment of one's performance.

The outcome-based (high-stakes testing) accountability inaugurated by NCLB incorporates **identifiability and evaluation** but gives short shrift to **reason giving and the mere presence of another** as less aggressive ways to foster accountability. In particular, policies and practices that increase transparency, making the relevant activity more visible to others, may evoke any or all of the first three accountability mechanisms even without imposing formal consequences. TEP Charter School's expectation that teachers observe each other routinely is one example of this; later, we describe other examples.

Accountability as Applied in Other Professions

Other professions—from law to engineering to architecture to medicine—typically use multiple forms of accountability that collectively make use of all four of the behavioral accountability mechanisms identified by Lerner and Tetlock.[5]

Many professions rely on *rule-based accountability,* which sets rules that delineate mandated or forbidden activities. Rule-based accountability relies on the identifiability of actors and sometimes includes an evaluation component. Rule-based accountability is common in K–12 education: State education codes, regulations, and teacher contracts create rule-based accountability.

Market-based accountability is also pervasive in most fields. Dentists and engineers, for example, are

accountable to their customers, via their consumer choice, in a way that public school educators typically are not. Under market-based accountability, clients or customers can hold professionals responsible by choosing among providers. Market-based accountability applies the identifiability mechanism, and it encourages providers to describe and explain their services, thereby invoking reason giving as well. Market-based accountability is relevant in private schools; in public charter schools; and, in a limited way, in conventional public schools, to the extent that families have the ability to move to a desirable school district.

In many fields, *professional accountability* systems go beyond what markets and government regulations require and often evoke all four behavioral mechanisms of accountability. Professional organizations impose standards for entry, provide resources for continuing learning, and set standards of practice that may be enforced by direct observation—for example, medical residents who are closely supervised by attending physicians. In contrast, although K–12 education includes some forms of professional accountability—for example, certification requirements—this sector historically has held only modest requirements for entering the profession and minimal requirements for maintaining professional status.

The kind of *outcome-based accountability* that has been the primary focus of education policy over the last two decades has not generally been common in other fields, perhaps because market-based accountability serves the same function. Tort law, in which plaintiffs'

attorneys are paid only if they win, is a notable exception. Outcome-based accountability is increasingly being attempted in health care,[7] for example, with financial penalties for high readmission rates of patients after hospital discharge.

In sum, many forms of accountability are used in various fields, and the different forms of accountability evoke different psychological mechanisms. Table 1 provides an overview of the different mechanisms evoked by different accountability types, with examples (outside of K–12 education) in each relevant cell.

These examples from other fields highlight the fact that policymakers have more tools available than just outcome-based accountability (high-stakes testing) to enhance school performance. Indeed, outcome-based accountability systems ignore two of the four behavioral mechanisms that promote accountability: mere presence of another and reason giving. In the rest of this article, we consider applications of all of these accountability types, using research from the field and the laboratory to inform ways that K–12 accountability regimes might be designed to improve educational outcomes.

Outcome-Based Accountability

Twenty-five years ago, outcome-based accountability was almost unknown in K–12 schooling. The education standards movement that gained steam during the 1990s promoted clear performance expectations for each grade level and tests to measure students' proficiency. Beginning in 1994, federal law (the Improving

Table 1. Accountability types in policy and psychological accountability mechanisms

Psychological accountability mechanism	Accountability types in policy			
	Outcome based	Rule based	Market based	Professional
Mere presence of another				Surgical operating room with nurses in attendance
Identifiability	*Consumer Reports,* Zagat	Minimum certification requirements (various professions)	Branding	Membership in a professional organization
Reason giving			Annual report to company stockholders	Medical rounds with explanation of treatment
Evaluation	Contingent fees for attorneys	Driver licensing test		Bar exam

America's Schools Act—the predecessor of NCLB) required states to set proficiency standards, assess students in multiple grades, and report school-specific results. The primary behavioral mechanism used was identifiability, in the form of public reporting of results. Eight years later, NCLB added explicit sanctions to schools falling short of proficiency targets, which added evaluation as a behavioral mechanism. More recently, the federal government pushed states to extend outcome-based accountability from schools to individual educators by evaluating teachers and principals in part on the basis of their student's achievement growth; this growth, or lack thereof, is measured yearly for each student, and then averaged across all students served by a particular teacher or principal.

These initiatives were not informed by the behavioral science literature, which finds few positive effects for outcome-based accountability.[5] Because it does not constrain decisionmakers with rules, outcome-based accountability might be more effective than other forms of accountability at promoting innovation,[8] but this has not been extensively studied. The effectiveness of outcome-based accountability can be undermined by the sunk-cost bias, which makes decisionmakers more likely to pursue action because of prior investments, even when the odds of success are low.[9] In addition, tangible rewards sometimes undermine intrinsic motivation.[10]

Behavioral studies also find that outcome-based accountability may impair decisionmaking by eliciting stress and negative emotions, increasing a decision's difficulty.[11,12] Perhaps because stress burdens cognition,[13,14] outcome accountability sometimes fails to increase the use of strategies that require substantial effort.[5]

Additionally, accountability regimes can be counterproductive when they are viewed as illegitimate.[5] Many teachers are suspicious of *value-added models* (VAMs) that aim to measure their contributions to student achievement. Improving on cruder outcome-based accountability regimes that rely on student achievement levels, VAMs account statistically for factors outside of the teacher's control, including students' demographic characteristics and (most important) prior achievement. In essence, VAMs measure how much better or worse a teacher's students are doing relative to how well the same students would have done if taught by an average teacher. Although well-designed VAMs can produce unbiased (fair) measures of teachers' contributions to student achievement growth,[15,16] suspicions on the part of practitioners could undermine their ability to promote performance improvements.

Moreover, even fair measures of educators' contributions to student achievement can be problematic in a high-stakes accountability system. Tests cannot capture all of the skills and knowledge that schools seek to impart. Some evidence suggests that instructional practices that raise test scores differ from those that promote students' effort and long-term goals.[17] High-stakes testing encourages "teaching to the test": Studies have found that many schools have narrowed their curriculums, focusing on reading and math to the exclusion of other subjects,[18] and spend a growing proportion of class time specifically preparing for the tests.[19,20] In extreme cases, educators have been caught cheating. Teacher-developed *student learning objectives,* increasingly used as outcome-based accountability measures, may be especially susceptible to inflation, because teachers grade themselves.[21] In short, test-based accountability may have the paradoxical effect of undermining the validity of the test itself.[22]

Even so, despite these unintended effects of outcome-based accountability, most existing field studies of the impact of high-stakes testing suggest positive effects in at least some schools, grades, and subjects.[23–27] (In the last cited reference, Deming et al. found mixed results in different schools.) The effects of performance-pay incentives for teachers on student achievement have been mixed, ranging from no effects to small positive effects.[28–30]

Rule-Based Accountability

Rule-based accountability relies on identifiability and sometimes on evaluation. Historically, rule-based accountability has been used to set constraints and conditions, such as in state lists of approved textbooks; contractual rules about working conditions, hours, and class size; and federal and state spending regulations. Through most of the past century, teachers had wide discretion about instruction,[31] which surely is one source of the wide variation in effectiveness in promoting student achievement.[32] Rules and protocols may have ensured minimum standards and reduced the most egregious inequities, but they may have also reduced opportunities for innovation. Charter schools were

created in part to allow innovation that would be less constrained by traditional rules, which may be particularly restrictive in the context of new instructional technologies that permit educators to organize schools and classrooms in novel ways.

Recently, some districts and school management organizations have become more directive about elements of instruction and school operations, pursuing a maximal version of rule-based accountability in which all teachers of the same courses may be expected to cover the same material at the same pace.[33] Principals have been asked to take on greater responsibilities as instructional leaders. Pacing guides are commonly used, and instruction is tied to state standards, with some lesson plans scripted to the minute.[34]

Limited evidence supports a maximal rule-based approach in schools. Scripted direct instruction has been found to promote student achievement in elementary grades.[35] But maximal rule-based accountability could become counterproductive, because psychological studies find that close monitoring often reduces intrinsic motivation.[36] The behavioral science literature also shows that intense monitoring can exacerbate the sunk-cost bias,[37] undermine innovation, and entrench suboptimal practices.[38] The perception of rules as illegitimate can produce a boomerang effect, leading people to react against the rules.[39] It is easy to imagine all of these effects operating in the classroom context, potentially undermining teaching and learning. Indeed, aggressive rule-based accountability may be especially unsuited to teaching, because it is an inherently complex task that requires daily adjustments and judgments. Highly scripted instructional programs might be useful in ensuring a minimal level of acceptable practice, but they are unlikely to promote excellence in teaching, particularly for deep and complex curricular material.

Market-Based Accountability

Market-based accountability is based on classical economic principles rather than newer behavioral approaches. It involves the identifiability and reason-giving mechanisms: Where choice is an option, schools chosen by families must be identifiable and attractive to parents. Historically, market-based accountability did not play a substantial role in U.S. public education. Operating alongside tuition-based private schools,

the public system has been based on the *common school model,* which assumes that each community will educate its children together, with school districts maintaining local monopolies on publicly supported education.[40,41]

Policymakers have shown increasing interest in incorporating market-based accountability into education, reasoning that local monopolies controlled by school boards may not produce the best schools[42,43] and that giving families choice in schooling is inherently valuable.[44] *Vouchers*—scholarships for tuition at private schools—have been advocated by conservatives (and a few liberals)[44] ever since Milton Friedman proposed the idea over 60 years ago,[42] and publicly funded voucher programs have been established in several states. Charter schools—publicly funded schools of choice that are open to all students, do not charge tuition, and operate outside of the direct control of local school districts[45]—represent a newer market-based approach and a new kind of public school, and they have received support across the political spectrum (as well as opposition from teachers unions and supporters of conventional public schools). The first charter schools opened only a quarter of a century ago; today, over 6,000 operate in more than 40 states.

Empirically, the evidence on the effects of market-based schools on student test scores and longer term educational attainment, although not definitive, suggests that the concept holds promise. In some contexts and locations, charter schools are producing substantial positive effects,[46–49] but their performance varies widely.[50,51] A few studies of small-scale voucher programs have found positive educational impacts, particularly for African-American students.[52,53] Louisiana's statewide voucher program for students in low-performing public schools, in contrast, has been shown to have substantial negative effects on student achievement in its first years of operation.[54,55]

K–12 schooling differs from other services in ways that might make exclusive reliance on the market suboptimal. First, the classic principal–agent problem—aligning the interests of clients and agents (educators)—is complicated by the involvement of multiple clients (students, parents, and the public), whose interests may not be fully aligned. Second, children are not fully capable of assessing their own best interests. In addition, students' educational experiences are affected not

only by school quality but by externalities, including characteristics of other students.[56] As a consequence, an unfettered market may produce segregated schools, as parents with high levels of knowledge, wealth, or motivation seek out schools that educate children from families like their own.

Relatedly, skeptics worry that market-based schools will drain conventional public schools of funding and motivated families. Supporters argue that breaking the local monopoly (that is, giving families publicly funded options other than those offered by the local school district) produces healthy competitive pressure that will benefit all students. The research base provides very little evidence that reallocation of students and resources produces any harm for students who remain in conventional public schools. Only one study has found a negative effect of charter schools on student achievement in nearby district-operated schools,[57] several studies have found no effects on students in nearby schools,[58–60] and a few studies have found positive effects.[60–62]

Another externality relates to the original rationale for public education: Society benefits from the inculcation of the knowledge, skills, and attitudes necessary for effective citizenship. Historically, this key rationale for the common school model implicitly assumed that effective education of citizens required public operation of schools.[39,40] The fact that the education of citizens is a public good argues against relying entirely on market mechanisms.

In fact, existing school choice programs rarely rely exclusively on market accountability. Charter schools are exempt from some forms of regulation but, like conventional public schools, are subject to rule-based and outcome-based accountability. Charter schools cannot charge tuition, their students must take the same high-stakes tests as students in conventional public schools, and these schools (typically) must admit all applicants, as space allows. Moreover, charter schools operate under the supervision of publicly empowered authorizers. Even the private schools that participate in voucher programs typically must submit to some regulation to receive public funds. Milwaukee's program, the longest-operating publicly funded voucher program in the country, imposes requirements for instructional time, forbids tuition, requires state assessments, and does not allow selective admissions.

Professional Accountability

Prominent voices are calling for greater professional accountability that would give K–12 teachers support, opportunities for collaboration, and training while also setting higher expectations.[2,3] Professional accountability in education can take many forms, involving all four of the behavioral accountability mechanisms. For example, licensing and professional reviews involve evaluation; observations and assistance by supervisors, instructional coaches, peers, or mentors involve identifiability, reason giving, and sometimes evaluation; collaboration and coteaching involve the mere presence of another and reason giving.

Traditional and Novel Versions of Professional Accountability

States have long applied requirements for teaching licenses, including coursework, student teaching, and passing exams. Teacher contracts generally reward master's degrees and experience as proxies for professional skill. But master's degrees have little or no relation to improved student achievement,[63,64] and most studies find professional development has no effect on student achievement.[65,66] Tellingly, traditional teacher evaluations have typically concluded that 98% to 99% of teachers are satisfactory, with tiny percentages falling short of satisfactory and no one better than satisfactory, because there was typically no rating category available to identify exemplary teachers.[67] Meanwhile, state laws and teacher contracts that award permanent tenure insulate teachers from professional accountability (as well as from other forms of accountability).

More robust and ambitious forms of professional accountability may hold more promise. Licensing and professional requirements at a high enough level—such as the certification process of the National Board for Professional Teaching Standards—might help identify especially effective teachers.[68] Because teacher quality is the most influential school-controlled factor in student achievement growth,[32] several initiatives of the Obama administration have promoted increased rigor in teacher evaluation, encouraging, for example, the use of multiple performance measures and multiple rating categories that include not only the traditional categories of *satisfactory* and *unsatisfactory* but also categories that

recognize high-performing teachers. States and school districts are adopting extensive rubrics for the observation and rating of teaching practice. There is a risk that the new systems could deteriorate into compliance exercises that resemble traditional rule-based accountability mechanisms, but if they function as intended, they could substantially improve practice. Robust professional accountability systems—unlike outcome-based, rule-based, and market-based accountability—include tools and resources to help teachers improve their skills. If taken beyond screening and compensation reform, they have the unique advantage of coupling accountability with support for improvement.

Novel forms of professional accountability might include new job descriptions and training methods. Some school districts have recently created teacher residency programs modeled on medical residencies, in which aspiring teachers spend much more time in the classroom during their training. Other districts and schools are giving teachers leadership opportunities, such as serving as instructional coaches who help their colleagues develop teaching skills.[69,70]

In most professions, professional accountability includes being answerable to clients.[71] K–12 schooling, in contrast, traditionally involves little direct accountability of educators to students. Nonetheless, school districts such as the Pittsburgh Public Schools are now including student surveys in new teacher evaluation systems. Some are using teacher surveys in principal evaluation,[72] applying the business world's 360-degree feedback to the academic workplace.

Another professional accountability system is an intensive review of school quality conducted by independent, expert educators, as is common in British schools. A school quality review involves an extended visit by outside experts who observe instruction; interview teachers, students, and parents; and examine school performance data. The review concludes with clear recommendations for improvement. Scholars such as Marc Tucker of the National Center on Education and the Economy,[2] Linda Darling-Hammond of Stanford,[3] and Helen Ladd of Duke[73] have proposed that school quality reviews be included as part of reformed accountability systems in the United States.

Professional accountability would make teaching more transparent, potentially activating all four behavioral accountability mechanisms. Indeed, rich professional accountability systems emphatically reject allowing teachers complete discretion in the classroom, under the assumption that there are standards of practice to which teachers should be held. Using transparency to promote professional accountability is a significant departure from an older, rule-based approach that values teacher autonomy over accountability. The old approach is regrettably evident in the rules of the Chicago Public Schools, which explicitly prohibit the use of classroom video recordings for evaluating teachers[74]—in sharp contrast to the expectations and transparent culture of the TEP Charter School where teachers observe each other every week. Schools like the Kauffman Charter School in Kansas City have gone one step further, literally making teaching transparent by giving classrooms interior windows that make them visible to adults in the hallways.

Greater teaching transparency is common in some countries that consistently outperform the United States in international comparisons of student achievement. A recent international study of educational practice found that although responding American teachers were more likely than their international peers to receive feedback from principals, only 11% received feedback from mentors, versus 39% in Japan, 38% in Singapore, and 24% in Australia.[75] Further, only 27% of responding American teachers received feedback from colleagues, versus 84% in Korea, 57% in the Netherlands, and 43% in Finland. All of those countries outperformed the United States in math, reading, and science in the most recent study of the Program for International Student Assessment.[76] American teachers were also far less likely than their counterparts in other countries to receive feedback from student surveys and less likely to report that the feedback they received led to public recognition, career advancement, or increased compensation.

Indeed, one study in an American urban school district found that improvements in student achievement were associated with teams of teachers who had strong mutual professional ties and with individual teachers who had strong ties with their principals.[77] Professional accountability could promote ties among teachers, potentially developing the social capital and trust that have been found to be markers of effective schools.[78]

Behavioral Evidence on Professional Accountability

Many studies in psychology demonstrate the favorable effects of requiring people to justify their decisions to

others, a common expectation of professional account-ability systems. One study found that requiring subjects to justify their decisions encourages high-effort strat-egies that are sensitive to evidence that can inform a decision, alleviating mistakes and inconsistencies.[79] Similarly, another study found that the need to justify decisions stimulated systematic thinking and attention to evidence.[80] In a third study, a justification requirement reduced reliance on stereotypic thinking.[81] Subjects who had to justify their judgments have also been found to be less likely to overattribute responsibility to individuals rather than situations.[82]

Other studies suggest that an increased sense of control—which might be promoted by a professional accountability environment that promotes initiative—may improve performance on attention-demanding tasks, promote more considerate decisionmaking, and assist memory formation.[83,84] More generally, profes-sional accountability may best encourage the system-atic, effortful, and self-critical thinking associated with even-handed, accurate reasoning.[85]

Professional accountability is also compatible with the behavioral nudges that are increasingly being adopted in various areas of public policy.[86] Field trials have demonstrated, for example, that appealing to social norms (using a wording like, "most people like you do X," following the work of Robert Cialdini)[87] powerfully influences behavior in contexts ranging from collecting taxes[88] to motivating parents to keep their children in school.[89] This suggests the possibility that providing relevant, appropriate evaluation feedback to teachers could lead to improvement even in the absence of explicit consequences.

Even though professional accountability is compat-ible with low-cost behavioral nudges, many forms of professional accountability are expensive or make substantial demands on educators. Teachers require time to observe each other and provide feedback. Instructional coaches need to be hired. School quality reviews must be staffed. More research is needed to assess whether some forms of professional account-ability are more cost-effective than others.

Field Evidence on Professional Accountability

A few studies have examined new forms of profes-sional accountability for educators. Teacher residencies are showing promise in producing high-performing teachers and keeping them in the classroom,[90] and early evidence on the effects of instructional coaching on student achievement is encouraging.[91–93] Several recently developed rubrics for observing and evalu-ating instructional practice have produced evaluation ratings that are correlated with teachers' contributions to student achievement,[94–96] suggesting that careful observation can produce feedback that could improve student outcomes.

In higher education, student evaluations of teachers have had positive effects on teaching.[97] Recent studies examining student surveys in K–12 schools have found the results are (modestly) correlated with measures of teachers' contributions to student achievement,[94,95] suggesting that they hold promise.

An intensive, peer-based teacher evaluation system used in Cincinnati offers encouraging evidence on formal, job-embedded professional accountability. Participating teachers substantially increased their effec-tiveness in raising student achievement during and after the year they were evaluated by peers—even though the evaluation criteria were based entirely on professional practice, not on test results.[98] Thomas Dee of Stan-ford University and James Wyckoff of the University of Virginia found that Washington, DC's ambitious teacher evaluation system—which uses multiple classroom observations by multiple observers, as well as measures of teachers' impacts on their students' achievement growth—produced positive effects on student achieve-ment.[99] And one experimental study found that nudging school principals with information about teachers' effectiveness (their prior contributions to student achievement) raised test scores and increased the attri-tion rate of low-performing teachers, even without incorporating the information in a formal, high-stakes evaluation measure.[100]

Increasing Professional Accountability and Transparency Using a Multimode Approach

The outcome-based accountability that has been the focus of policymakers' attention has produced some positive results, but relying on it exclusively is unlikely to produce large, sustained improvements and can lead to unintended and undesirable side effects. The evidence from behavioral science laboratories and from the field makes clear that other accountability approaches can also produce favorable results, suggesting that a

narrow focus on outcome-based accountability leaves important tools unused.

Below is a modified version of the table from the beginning of the article. As in the original table, in Table 2 we provide examples illustrating how different accountability types can evoke different psychological accountability mechanisms. Here the examples are specific to K–12 education, confirming that policymakers and educators have a wide range of accountability tools to use, just as they do in other professions and fields.

After reviewing the evidence from behavioral science and the field literature, we propose that policymakers designing K–12 accountability systems should consider three key points. *First, policymakers should make greater use of professional accountability, which has historically been underutilized in education.* As Table 2 indicates, different professional approaches can invoke all four of the motivational mechanisms inherent in different types of accountability. In the past, K–12 schooling has used professional accountability in limited ways, largely related to establishing minimum standards for entry or promotion. Forms of professional accountability that are more robust and ambitious could set higher expectations for professional practice and simultaneously create feedback mechanisms that help educators improve their practice.

Second, transparency alone can create professional accountability. All of the examples of professional accountability included in Table 2 involve making educational practice more transparent to other educators. Transparency of practice activates several behavioral accountability mechanisms (mere presence of another, identifiability, and reason giving) that powerfully influence behavior. Transparency also provides an opportunity to offer feedback to improve performance. Professional accountability can involve transparency in various ways, including not only the literal transparency of glass-walled classrooms but also peer observation and evaluation, instructional coaching, and 360-type feedback.

Third, multiple forms of accountability can be used in complementary ways, creating a complete system that encourages and supports the continuous improvement of educational practice and outcomes. Using multiple approaches can play to the advantages of each type while minimizing disadvantages—much as TEP Charter School uses outcome-based accountability (in the form of bonuses for school-wide performance) and market-based accountability (it must attract students to survive) alongside a rich professional accountability system. For example, Ken Frank of Michigan State University has proposed a form of school governance that would utilize multiple modes of accountability, making the principal accountable to a community board and giving the principal greater authority to remove the lowest performing teachers while at the same time giving teachers more say in school operations, including the evaluation of the principal.[101] Shefali Patil of the University of Texas, Ferdinand Vieider of Reading University (UK), and Philip Tetlock of the University of Pennsylvania recently noted that outcome-based accountability

Table 2. Accountability types and psychological accountability mechanisms with applications in kindergarten through 12th grade schooling

Psychological accountability mechanisms	Accountability types			
	Outcome based	Rule based	Market based	Professional
Mere presence of another				Classroom windows
Identifiability	Public reporting of school-wide test results	Minimum certification requirements	School choice	Peer observation
Reason giving			Charter-school enrollment fairs	Instructional coaching
Evaluation	Value-added model incorporated in teacher evaluations	Formal observation by principal		Peer review; advanced certification

may better promote innovation, but process-based accountability (including forms of professional accountability) may better promote the use of best practices.[38] They suggested that the disadvantages of both types might be counteracted by systems that promote the empowerment of decisionmakers to rethink ineffective practices, encourage focus on outcomes, and facilitate organizational learning.[102] This kind of empowerment is implicit in professional accountability, and it can also be incorporated into an outcome-based accountability system that communicates a desire to achieve shared objectives. Similarly, according to a National Academies report on high-stakes testing,[6] external rewards are most likely to be effective when they are closely aligned with educators' intrinsic aims, promoting "autonomous motivation."

Moreover, organizational learning requires feedback for improvement.[102] Outcome-based accountability and market-based accountability can create incentives for improved performance, but they provide no information or resources to help professionals actually improve their performance. In a more integrated approach, transparency of practice creates opportunities for educators to improve, rich data on student outcomes help diagnose students' needs, and rewards for success encourage educators to innovate in productive ways.

Much of the backlash against high-stakes testing has come in the context of teacher evaluations that include student achievement growth as a formal component. There are good reasons to avoid relying exclusively on test scores for accountability. But the last decade has seen an enormous amount of work on other elements of the evaluation system that are related to professional accountability, including higher-quality classroom observations, an increase in the number of rating categories to differentiate teachers at the high end of the performance distribution as well as the low end, and the inclusion of student feedback. Efforts to roll back the use of test scores in educator evaluation risk undermining these fledgling efforts to promote increased professional accountability. In Los Angeles, for example, the district and teachers union recently agreed to new evaluation policies that eliminated the highest rating category, with the result that it is impossible for a teacher to exceed expectations.[103] Refusing to recognize exemplary performance among teachers is not only contrary to the evidence about variation in teacher effectiveness but also incompatible with the goal of creating a system of continuous improvement in professional practice.

As policymakers and educators take advantage of the Every Student Succeeds Act's new flexibility, we hope they recognize that reducing a near-exclusive reliance on outcome-based accountability does not have to mean reducing accountability as a whole; that a wide range of tools are available for creating a richer accountability system that can promote continuous improvement; and that professional accountability should play an important role in that system, raising expectations for teachers and schools while providing better opportunities to meet those raised expectations.

author affiliation

Gill, Mathematica Policy Research; Lerner, Harvard Kennedy School, Harvard University; Meosky, Harvard Kennedy School, Harvard University. Corresponding author's e-mail address: bgill@mathematica-mpr.com

References

1. Furgeson, J., McCullough, M., Wolfendale, C., & Gill, B. (2014). *The Equity Project Charter School: Impacts on student achievement*. Cambridge, MA: Mathematica Policy Research.
2. Tucker, M. S. (2014). *Fixing our national accountability system*. Washington, DC: National Center on Education and the Economy.
3. Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives, 22*(86).
4. Center on Reinventing Public Education. (2014). *Designing the next generation of state education accountability systems: Results of a working meeting*. Retrieved from http://www.crpe.org/sites/default/files/CRPE_designing-next-gen-state-ed-accountability.pdf
5. Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255–275.
6. Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. Washington, DC: National Academies Press.
7. Gold, M. (2010). *Accountable care organizations: Will they deliver?* Princeton, NJ: Mathematica Policy Research.

8. Patil, S. V., & Tetlock, P. E. (2014). Punctuated incongruity: A new approach to managing trade-offs between conformity and deviation. *Research in Organizational Behavior, 34,* 155–171.

9. Simonson, I., & Staw, B. M. (1992). De-escalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology, 77,* 419–426.

10. Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125,* 627–688.

11. Zhang, Y., & Mittal, V. (2005). Decision difficulty: Effects of procedural and outcome accountability. *Journal of Consumer Research, 32,* 465–472.

12. Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes, 1,* 1–17.

13. Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science, 65,* 221–244.

14. Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York, NY: Springer.

15. Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation.

16. Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104,* 2593–2632.

17. Ferguson, R. F. (with Danielson, C.). (2014). How Framework for Teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 98–143). Hoboken, NJ: Jossey-Bass.

18. Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effect of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35,* 252–279.

19. Hamilton, L. S., Stecher, B. M., & Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. *Education Inquiry, 3,* 149–170.

20. Jennings, J., & Sohn, H. (2014). Measure for measure: How proficiency-based accountability systems affect inequality in academic achievement. *Sociology of Education, 87,* 125–141.

21. Gill, B., English, B., Furgeson, J., & McCullough, M. (2014). *Alternative student growth measures for teacher evaluation: Profiles of early-adopting districts.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

22. Campbell, D. T. (1976). *Assessing the impact of planned social change* (Occasional Paper No. 8). Hanover, NH: Dartmouth College Public Affairs Center.

23. Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30,* 418–446.

24. Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24,* 305–331.

25. Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina* (NBER Working Paper No. 20511). Cambridge, MA: National Bureau of Economic Research.

26. Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics, 93,* 1045–1057.

27. Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2013). *School accountability, postsecondary attainment, and earnings* (NBER Working Paper No. 19444). Cambridge, MA: National Bureau of Economic Research.

28. Chiang, H., Wellington, A., Hallgren, K., Speroni, C., Herrmann, M., Glazerman, S., & Constantine, J. (2015). *Evaluation of the Teacher Incentive Fund: Implementation and impacts of pay-for-performance after two years.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

29. Springer, M. G., Pane, J. F., Le, V., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. (2012). Team pay for performance: Experimental evidence from the Round Rock Pilot Project on Team Incentives. *Educational Evaluation and Policy Analysis, 34,* 367–390.

30. Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years.* Washington, DC: Mathematica Policy Research.

31. Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

32. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417–458.

33. Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement.* New York, NY: MDRC.

34. Beatty, B. (2011). The dilemma of scripted instruction: Comparing teacher autonomy, fidelity, and resistance in the Froebelian kindergarten, Montessori, Direct Instruction, and Success for All. *Teachers College Record, 113,* 395–430.

35. Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 125–230.

36. Enzle, M. E., & Anderson, S. C. (1993). Surveillant intentions and intrinsic motivation. *Journal of Personality and Social Psychology, 64,* 257–266.

37. Ross, J., & Staw, B. M. (1993). Organizational escalation and exit: Lessons from the Shoreham nuclear power plant. *Academy of Management Journal, 36,* 701–732.

38. Patil, S., Vieider, F., & Tetlock, P. E. (2012). Process versus outcome accountability. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *Oxford handbook of public accountability* (pp. 69–89). New York, NY: Oxford University Press.

39. Baer, R., Hinkle, S., Smith, K., & Fenton, M. (1980). Reactance as a function of actual versus projected autonomy. *Journal of Personality and Social Psychology, 35,* 416–422.

40. Tyack, D., & Hansot, E. (1982). *Managers of virtue: Public school leadership in America, 1820–1980.* New York, NY: Basic Books.

41. Glenn, C. L., Jr. (1988). *The myth of the common school.* Amherst: University of Massachusetts Press.

42. Friedman, M. (1955). The role of government in education. In R. A. Solo (Ed.), *Economics and the public interest.* Piscataway, NJ: Rutgers University Press.

43. Chubb, J., & Moe, T. (1990). *Politics, markets, and America's schools.* Washington, DC: Brookings Institution Press.

44. Coons, J. E. (1992). School choice as simple justice. *First Things, 22,* 193–200.

45. Gill, B., Timpane, P. M., Ross, K. E., Brewer, D. J., & Booker, K. (2007). *Rhetoric versus reality: What we know and what we need to know about vouchers and charter schools.* Santa Monica, CA: RAND.

46. Abdulkadiroglu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics, 126,* 699–748.

47. Booker, K., Sass, T. R., Gill, B., & Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics, 29,* 377–415.

48. Dobbie, W., & Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy, 123,* 985–1037.

49. Gleason, P. M., Tuttle, C. C., Gill, B., Nichols-Barrer, I., & Teh, B. (2014). Do KIPP schools boost student achievement? *Education Finance and Policy, 9,* 36–58.

50. Gill, B., & Nichols-Barrer, I. (2014). Charter schools. In D. Brewer & L. Picus (Eds.), *Encyclopedia of education economics and finance* Thousand Oaks, CA: Sage.

51. Center for Research on Education Outcomes. (2013). *National charter school study.* Stanford, CA: Stanford University.

52. Wolf, P., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management, 32,* 246–270.

53. Chingos, M. M., & Peterson, P. E. (2015). Experimentally estimated impacts of school vouchers on college enrollment and degree attainment. *Journal of Public Economics, 122,* 1–12.

54. Mills, J. N., & Wolf, P. J. (2016). *The effects of the Louisiana Scholarship Program on student achievement after two years* (Louisiana Scholarship Program Evaluation Report 1). Fayetteville: University of Arkansas, School Choice Demonstration Project.

55. Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2015). School vouchers and student achievement: Evidence from the Louisiana Scholarship Program (NBER Working Paper No. 21839). Cambridge, MA: National Bureau of Economic Research.

56. Zimmer, R. W., & Toma, E. F. (2000). Peer effects in public and private schools across countries. *Journal of Policy Analysis and Management, 19,* 75–92.

57. Imberman, S. (2011). The effect of charter schools on achievement and behavior of public school students. *Journal of Public Economics, 95,* 850–863.

58. Bifulco, R., & Ladd, H. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1,* 50–90.

59. Bettinger, E. P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review, 24,* 133–147.

60. Zimmer, R., Gill, B., Booker, T. K., Lavertu, S., Sass, T. R., & Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition.* Santa Monica, CA: RAND.

61. Winters, M. A. (2012). Measuring the competitive effect of charter schools on public school student achievement in an urban environment: Evidence from New York City. *Economics of Education Review, 31,* 293–301.

62. Jinnai, Y. (2014). Direct and indirect impact of charter schools' entry on traditional public schools: New evidence from North Carolina. *Economics Letters, 124,* 452–456.

63. Chingos, M. M., & Peterson, P. E. (2010). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30,* 449–465.

64. Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26,* 673–682.

65. Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38,* 915–945.

66. Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 151–180). San Francisco, CA: Jossey-Bass.

67. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* New York, NY: New Teacher Project.

68. Cowan, J., & Goldhaber, D. (2015). National Board certification and teacher effectiveness: Evidence from Washington (CEDR Working Paper 2015-3). Seattle: University of Washington Bothell, Center for Education Data and Research.

69. Barnwell, P. (2015, February 18). Why schools need more 'hybrid' teaching roles. *Education Week.* Retrieved from http://www.edweek.org/tm/articles/2015/02/18/why-schools-need-more-hybrid-teaching-roles.html

70. Gawande, A. (2011, October 3). Personal best. *The New Yorker.* Retrieved from http://www.newyorker.com/magazine/2011/10/03/personal-best

71. Newton, L. H., Hodges, L., & Keith, S. (2013). Accountability in the professions: Accountability in journalism. *Journal of Mass Media Ethics, 19,* 166–190.

72. Porter, A. C., Polikoff, M. S., Goldring, E. B., Murphy, J., Elliott, S. N., & May, H. (2010). Investigating the validity and reliability of the Vanderbilt Assessment of Leadership in Education. *The Elementary School Journal, 111,* 282–313.

73. Ladd, H. F. (2016, May 26). Now is the time to experiment with school inspections for accountability [Blog post]. Retrieved from https://www.brookings.edu/2016/05/26/now-is-the-time-to-experiment-with-inspections-for-school-accountability/

74. Chicago Public Schools. (2014). *REACH Students educator evaluation handbook 2014–15.* Chicago: Author.

75. Organisation for Economic Co-operation and Development. (2014). *TALIS 2013 results: An international perspective on teaching and learning.* Paris, France: Author.

76. Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results in focus.* Paris, France: Author.

77. Pil, F. K., & Leana, C. (2009). Applying organizational research to school reform: The effects of human and social capital on student performance. *Academy of Management Journal, 52,* 1101–1124.

78. Bryk, A. S., & Schneider, B. (2004). *Trust in schools: A core resource for improvement.* New York, NY: Russell Sage Foundation.

79. Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes, 52,* 292–306.

80. Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24,* 563–574.

81. Bodenhausen, G. V., Kramer, G. P., & Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66,* 621–632.

82. Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

83. Sherman, G. D., Lee, J. J., Cuddy, A. J. C., Renshon, J., Oveis, C., Gross, J. J., & Lerner, J. S. (2012). Leadership is associated with lower levels of stress. *Proceedings of the National Academy of Sciences, USA, 109,* 17903–17907.

84. Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

85. Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Farrar, Straus and Giroux.

86. Thaler, R. H., & Sunstein, C. B. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

87. Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika, 72,* 263–268.

88. Hallsworth M., List, J., Metcalfe, R., & Vlaev, I. (2014). *The behavioralist as tax collector: Increasing tax compliance through natural field experiments* (NBER Working Paper No. 20007). Cambridge, MA: National Bureau of Economic Research.

89. Kraft, M., & Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review, 47,* 49–63.

90. Hallberg, K., & Green, G. (2015, March 11). How can we hire and keep high quality teachers in struggling schools? [Blog post]. Retrieved from http://educationpolicy.air.org/blog/how-can-we-hire-and-keep-high-quality-teachers-struggling-schools

91. Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., & Lake, R. (2012). *Charter-school management organizations: Diverse strategies and diverse student impacts.* Cambridge, MA: Mathematica Policy Research.

92. Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making. *Educational Policy, 24,* 872–907.

93. Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis, 37,* 542–566.

94. Kane, T. J. (2012). Capturing the dimensions of effective teaching. *Education Next, 12*(4), 35–41.

95. Chaplin, D., Gill, B., Thompkins, A., & Miller, H. (2014). *Professional practice, student surveys, and value added: Multiple measures of teacher effectiveness in the Pittsburgh Public Schools* (REL 2014-024). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

96. Walsh, E., & Lipscomb, S. (2013). *Classroom observations from Phase 2 of the Pennsylvania Teacher Evaluation Pilot: Assessing internal consistency, score variation, and relationships with value added.* Cambridge, MA: Mathematica Policy Research.

97. L'Hommedieu, R., Menges, R. J., & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology, 82,* 232–241.

98. Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102,* 3628–3651.

99. Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34,* 267–297.

100. Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review, 102,* 3184–3213.

101. Frank, K. (2012, February 24). Constitution for effective school governance [ID No. 16715]. *Teachers College Record.* Available from http://www.tcrecord.org

102. Schillemans, T., & Smulders, R. (2015). Learning from accountability? Whether, what, and when. *Public Performance & Management Review, 39,* 248–271.

103. Blume, H. (2016, June 13). Less test-iness over L.A. teacher evaluations. *Los Angeles Times.* Retrieved from http://www.latimes.com/

Healthy through habit: Interventions for initiating & maintaining health behavior change

Wendy Wood & David T. Neal

Summary.   Interventions to change health behaviors have had limited success to date at establishing enduring healthy lifestyle habits. Despite successfully increasing people's knowledge and favorable intentions to adopt healthy behaviors, interventions typically induce only short-term behavior changes. Thus, most weight loss is temporary, and stepped-up exercise regimens soon fade. Few health behavior change interventions have been successful in the longer term. In this article, we unpack the behavioral science of health-habit interventions. We outline habit-forming approaches to promote the repetition of healthy behaviors, along with habit-breaking approaches to disrupt unhealthy patterns. We show that this two-pronged approach—breaking existing unhealthy habits while simultaneously promoting and establishing healthful ones—is best for long-term beneficial results. Through specific examples, we identify multiple intervention components for health policymakers to use as a framework to bring about lasting behavioral public health benefits.

In 1991, the National Cancer Institute and industry partners rolled out a nationwide educational public health   campaign—the 5 A Day for Better Health Program—to boost consumption of fruits and vege-tables. The campaign was remarkably successful in changing people's knowledge about what they should eat: Initially, only 7% of the U.S. population understood that they should eat at least five servings of fruit and vegetables per day, whereas by 1997, fully 20% were aware of this recommendation.[1] Unfortunately, actual fruit and vegetable consumption remained flat. During the years 1988 to 1994, 11% of U.S. adults met this target amount of fruit and vegetable consumption, and the percentage did not shift during 1995–2002.[2] Another national campaign launched in 2007, called Fruit & Veggies—More Matters, also failed to move the fruit and vegetable consumption needle.[3]

These failures are not surprising. A body of research shows that many public health campaigns do success-fully educate and motivate people, especially in the short run. However, when push comes to shove, they often fail at changing actual behaviors and long-term

health habits, such as the consumption of optimal amounts of fruit and vegetables.[4,5]

Not all behavior change interventions fail to change behavior. Often, some behavior change happens, but it does not maintain over time.[6] To show how this works, we depicted the results of some of the highest quality health interventions to date in Figure 1. These studies all appeared in top scientific journals, used exemplary methods, and conscientiously assessed long-term success rates.[7–10] It is easy to see that most participants in these interventions got healthier in the short term (as shown by the initially increasing lines). They lost weight, exercised more, and gave up smoking. However, once the intervention ended, old patterns reemerged, and the new, healthy behaviors clearly waned over time (as shown by the eventually decreasing lines). The overall trajectory of behavior change can be described as a *triangular relapse pattern*.

It is tempting to believe that the failures in maintaining healthy behaviors depicted in Figure 1 are

Figure 1. The triangular relapse pattern in health behavior change over time

In these triangular relapse patterns, an initial spike in healthful behaviors during the intervention is followed by a decline following intervention back toward baseline. Panels A–D show four examples of behavior change interventions following this pattern for (A) weight loss, (B) gym visits, (C) quitting smoking, and (D) exercise. Mos = months; MVP = moderate to vigorous physical activity.

A. Intervention of financial incentives for weight loss
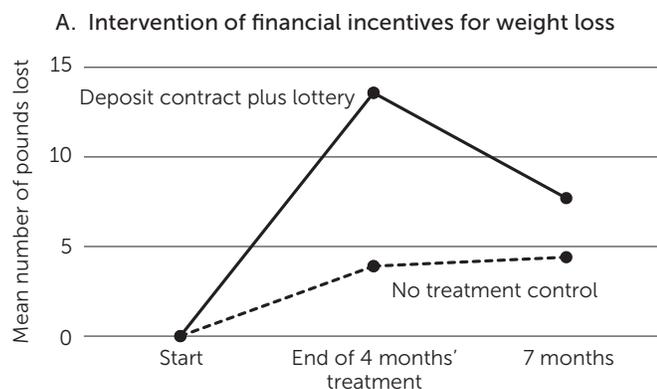
A: Mean pounds lost following a 4-month intervention of financial incentives for weight loss and after 3 months of no treatment (*N* = 57). Data are from "Financial Incentive–Based Approaches for Weight Loss: A Randomized Trial," by K. G. Volpp, L. K. John, A. B. Troxel, L. Norton, J. Fassbender, and G. Loewenstein, 2008, *Journal of the American Medical Association,* 300, p. 2635. Copyright 2008 by the American Medical Association.

B. Intervention of payment for gym visits
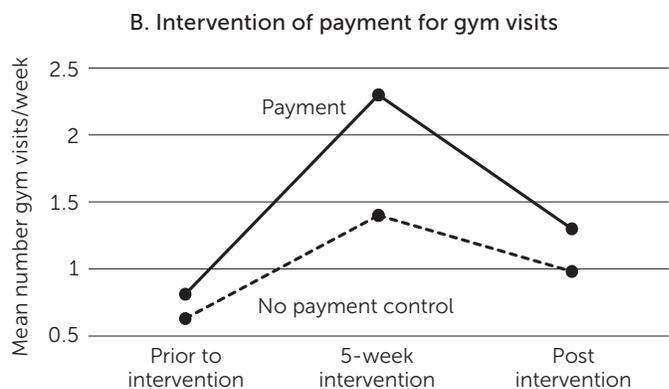
B: Mean gym visits per week prior to study (weeks -16 to -2), during 5 intervention weeks of payment for attending, and during 15 no-treatment weeks (weeks 6–21, *N* = 99). Data are from "Incentives to Exercise," by G. Charness and U. Gneezy, 2009, *Econometrica,* 77, p. 921, Figure 2b. Copyright 2009 by Wiley.

C. Intervention of smoking information and financial incentives to quit
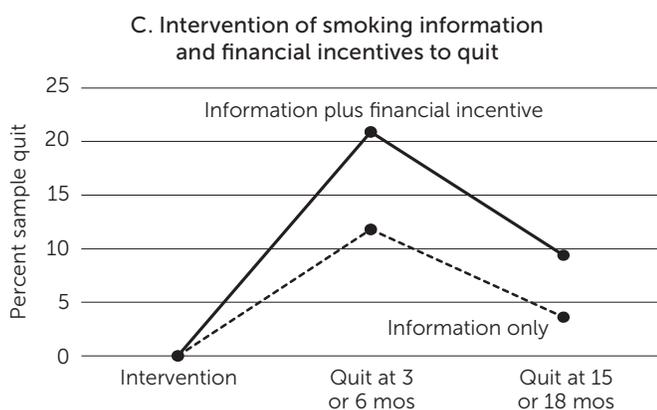
C: Percentage of participants who quit smoking (biochemically verified) at 3 or 6 months and at 15 or 18 months following intervention of information about smoking cessation programs paired with financial incentives (*N* = 878). Data are from "A Randomized, Controlled Trial of Financial Incentives for Smoking Cessation," by K. G. Volpp, A. B. Troxel, M. V. Pauly, H. A. Glick, A. Puig, D. A. Asch, . . . J. Audrain-McGovern, 2009, *New England Journal of Medicine,* 360, p. 703, Table 2. Copyright 2009 by the Massachusetts Medical Society.

D. Intervention via computer to encourage physical activity
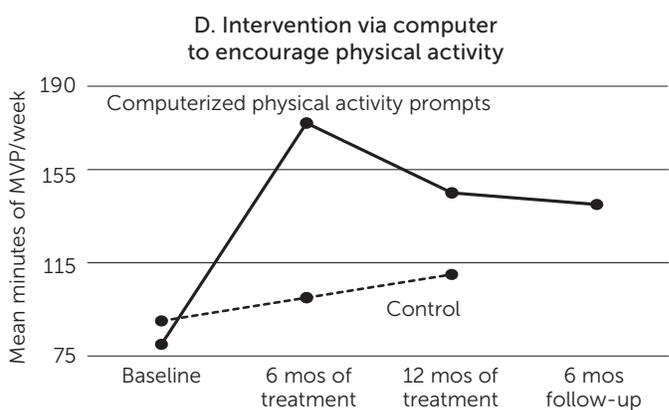
D: Mean number of minutes per week of moderate to vigorous physical exercise during computer-delivered interventions or health program controls at 6 months of treatment, 12 months of treatment, and 6 months after end of treatment (*N*s = 70 control and 75 computerized treatment at baseline; *N* = 61 computerized treatment at 18 months). Data are from "Exercise Advice by Humans Versus Computers: Maintenance Effects at 18 Months," by A. C. King, E. B. Hekler, C. M. Castro, M. P. Buman, B. H. Marcus, R. H. Friedman, and M. A. Napolitano, 2014, *Health Psychology,* 33, p. 195, Figure 1. Copyright 2014 by the American Psychological Association.

simply due to people's limited willpower. Surely many people struggle to inhibit the short-term gratifications of fast food and the lure of excessive TV watching and do not make the effort to stick to a balanced diet and regular exercise. However, relapse is not inevitable if behavior change interventions form healthy habits. In fact, research shows that people who consistently act in healthy ways in daily life do so out of habit. With heathy diet and exercise habits, they do not need to struggle with internal urges to act in unhealthy ways.[11,12] Another insight comes from the success of policy changes and health interventions in the last few decades that drastically reduced smoking rates in the United States. Antismoking campaigns have many components,[13] but the most successful parts targeted cigarette purchase and smoking habits as opposed to people's willpower and self-control. In this article, we use these insights as a framework to construct interventions that break unhealthy habits and encourage the adoption of beneficial ones.

Both breaking and creating habits are central to behavior change. Habits play a significant role in people's failure to adopt and stick with what is best for their health. Eating habits are especially striking. Research has shown that people habitually consume food that they neither want nor even like.[14] For example, movie theater patrons with strong popcorn-eating habits consumed just as much stale, week-old popcorn as they did fresh popcorn, despite reporting that they hated the stale food.[15]

Fortunately, just as bad habits impede behavior change, good habits can promote it. As noted above, good habits ensure that people continue to act in healthy ways without constant struggle. For example, chocolate lovers who had formed a habit to eat carrots continued to make the healthy carrot choice even when chocolate became available.[16]

> *Existing habits are a significant impediment to people adopting and sticking with healthy behavior*

Habits represent context–response associations in memory that develop as people repeat behaviors in daily life. For example, after repeatedly eating hamburgers and pizza for dinner, a person is likely to find that dinnertime cues such as driving home from work and watching the evening news automatically activate thoughts of these foods and not vegetables.[17]

From a habit perspective, behavior change interventions are likely to fail unless they account for the ways in which people form healthy habits and break unhealthy ones. Although the research literature on behavior change offers sophisticated understanding of many intervention features (for example, offering appropriate incentives, tailoring messages to specific subsets of the target audience, tracking nonintrusive outcomes such as credit card charges), little attention has been paid to the importance of habits in maintaining lifestyle choices.

In the first part of this article, we explain how interventions create healthy habits. Essentially, healthy habit creation involves repeated performance of rewarding actions in stable contexts. The second part of the article addresses how interventions can break unhealthy habits by neutralizing the cues that automatically trigger these responses. Our set of habit-based interventions thus augments existing tools to promote automated performance of desired over undesired responses. Among existing tools, people are most likely to make a good choice when decisions are structured to make that choice easy,[18,19] when other people are making the same choice,[20,21] and after forming if-then plans.[22,23] Finally, we explain how habit-based interventions can be incorporated into health policies.

Promoting the Formation of New Habits

The three central components of habit formation are (a) behavioral repetition, (b) associated context cues, and (c) rewards (see Table 1).

Behavior change interventions form habits by getting people to act in consistent ways that can be repeated frequently with little thought. Habits develop gradually through experience, as people repeat a rewarded action in a stable place, time, or other context. Through repetition, the context becomes a sort of shorthand cue for what behavior will be rewarded in that context. People's habits essentially recreate what has worked for them in the past. In this way, habits lock people into a cycle of automatic repetition.

Once a habit has formed, it tends to guide behavior even when people might have intended to do something else.[24] Essentially, habits come to guide behavior instead of intentions. Early in habit formation, people might intentionally decide how to respond to achieve

Table 1. Three main components of habit formation interventions and examples of implementation in practice

Principle	Examples in practice
Frequent repetition	• School hand-washing interventions that involve practicing actual washing behavior in the restroom
Recurring contexts and associated context cues	• Public health campaigns linking changing smoke detector batteries to the start and end of daylight savings time • Medical compliance communications that piggyback medications onto existing habits such as mealtime
Intermittent rewards	• Free public transit days scheduled randomly • Coupons and discounts for fresh fruits and vegetables provided on an intermittent or random basis

a certain outcome. However, once a habit gains strength, people tend to habitually respond, for better or worse.[25] According to a study in the *British Journal of Health Psychology,* eating habits were stronger determinants of food choices than intentions or even sensitivity to food temptations.[26] When habits are healthy, outsourcing behavioral control to the environment in this way is beneficial. People keep on track by responding habitually when distractions, stress, and dips in willpower impede decision-making.[27] However, when habits are unhealthy, the automatic or environmental control of behavior impedes health and can create a self-control dilemma.

Next, we expand on the central components of habit formation and later address unhealthy habits.

The Three Central Habit-Forming Interventions

Behavior Repetition

Habit formation interventions create opportunities for and encourage frequent repetition of specific responses, but there is no single formula for success. In one study, participants chose a new health behavior to perform once a day in the same context (for example, eating fruit after dinner).[28] For some behaviors and some people, only 18 days of repetition were required for the behavior to become sufficiently automatic to be

performed without thinking. For other behaviors and participants, however, over 200 days of repetition were needed. Another study published in *Health Psychology*[29] found that people required 5 to 6 weeks of regular gym workouts to establish new exercise habits.[30]

Interventions may encourage repetition by visually depicting the physical act of repeating the desired behavior—think of the famous Nike advertisements advising, "Just Do It," while showing famous athletes and others engaged in vigorous exercise. Interventions in schools and other controlled environments could direct physical practice of the new habit by, for example, conducting hand-washing drills in bathrooms instead of merely teaching hygiene benefits and setting performance goals.[31] Hospitals and restaurants can similarly benefit from employees rehearsing best sanitation practices.

Longer interventions with frequent repetitions (vs. shorter interventions, with fewer repetitions) tend to be most successful because they are most likely to lead to the formation of strong habits. Such a pattern could explain the greater success of long-duration weight loss interventions.[5] Intervention length also might explain one of the most successful behavioral interventions: Opower's multiyear energy conservation programs.[32] These multicomponent interventions, involving smart meters and feedback about power use, have proved especially successful at limiting energy use, presumably because the extended intervention allowed consumers to form energy-saving habits.

Context Matters: Cues Trigger Habit Formation

Successful habit learning depends not only on repetition but also on the presence of stable context cues. Context cues can include times of day, locations, prior actions in a sequence, or even the presence of other people (see Table 1). Illustrating the importance of stable cues, almost 90% of regular exercisers in one study had a location or time cue to exercise, and exercising was more automatic for those who were cued by a particular location, such as running on the beach.[33] Other research shows that older adults are more compliant with their drug regimens when pill taking is done in a particular context in their home (for example, in the bathroom) or integrated into a daily activity routine.[34]

Implementation plans. Intervention programs to form healthy habits can promote stable habit cues in

several ways. People can be encouraged to create plans, or *implementation intentions*, to perform a behavior in a given context (for example, "I will floss in the bathroom after brushing my teeth").[18]

Forming implementation plans increases the likelihood that people will carry out their intentions.[35] Accordingly, these plans promote performance only for people who already intend to perform the healthy behavior (for example, people who want to floss more regularly),[36] and the efficacy of the intervention fades if their intentions change. Even so, implementation intentions may be a useful stepping stone on the path to creating habits because, as people act repeatedly on such intentions in a stable context, behavior may gradually become less dependent on intentions and gel into habits.

Piggybacking. Intervention programs also create cues by *piggybacking*, or tying a new healthy behavior to an existing habit. The habitual response can then serve as a cue to trigger performance of the new behavior. For example, dental-flossing habits were established most successfully when people practiced flossing immediately after they brushed their teeth, rather than before.[37] The large number of habits in people's daily lives provides many opportunities to connect a new behavior to an existing habit.[38] Successful examples include public information campaigns that link the replacement of smoke alarm batteries to another periodic activity— changing the clock for daylight savings; and medical compliance is boosted when a prescribed health practice (for example, taking pills) is paired with a daily habit (for example, eating a meal, going to bed).[39]

Rewards Promote Habit Formation

People tend to repeat behaviors that produce positive consequences or reduce negative ones (see Table 1). Positive consequences include the intrinsic payoff of a behavior, for instance, the taste of a sweet dessert or the feeling of accomplishment that comes from effectively meeting health goals.[40] Positive consequences also include extrinsic rewards, such as monetary incentives or others' approval. Avoiding negative consequences is illustrated by *contingency contracts*, such as when people agree to pay money for every swear word they utter or experience other negative consequences for failing to meet a goal.[41]

Habits form most readily when specific behaviors are rewarded. Especially during the initial stages of habit formation, specific incentives can increase people's motivation to do things they might typically avoid, such as exercising or giving up ice cream. In this sense, rewards can offset the loss of enjoyable activities in order to start a healthful behavior.

Other rewards are less successful at habit formation because they are too broad to promote specific habits. Overly general rewards include symbolic trophies, prizes that recognize strong performance, or temporal landmarks such as birthdays or the kickoff of a new calendar year. Only rewards that promote the repetition of specific actions contribute to habit formation.
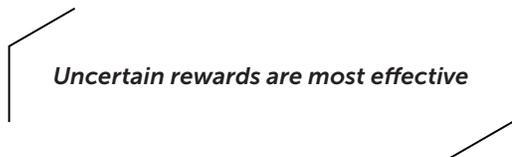
Uncertain rewards are most effective

Many decades of laboratory research have shown what kinds of rewards are most likely to motivate habits. Surprisingly, habits form best when rewards are powerful enough to motivate behavior but are uncertain in the sense that they do not always occur.[42] Uncertain rewards powerfully motivate repetition and habit formation. In learning theory terminology, such rewards are given on random-interval schedules.

Slot machines are a good example of uncertain rewards. People keep paying money into the machines because sometimes they win, sometimes they don't. This reward system is so powerful that slot machines are sometimes described as the crack cocaine of gambling. E-mail and social networking sites have similar effects: people keep checking on them because sometimes they are rewarded with interesting communications, but other times they get only junk. The key is that rewards are received probabilistically, meaning not for every behavior.

To date, few health interventions have used uncertain rewards.[43] Instead, most health interventions offer consistent, predicable rewards, such as payments received each time program participants go to the gym. Such rewards effectively drive short-term behavior changes, but they do not establish habits. When the rewards stop, people usually quit the behavior.[6] In part, people quit because predictable rewards can signal that a behavior is difficult, undesirable, and not worth performing without the reward.[44]

Behavior change interventions should give rewards in the way a slot machine does—at uncertain intervals

but often enough to sufficiently motivate people to perform the target healthy behavior. For example, discounts on fresh fruits and vegetables at grocery stores can be provided intermittently to encourage habitual produce purchases. The structure and routines of school and work environments are particularly well suited to providing uncertain rewards. School policies, especially in elementary schools, could be structured to provide occasional monitoring and reinforcements for healthy behaviors such as hand washing after using the restroom or fruit and vegetable consumption during school lunches.

The Three Main Habit-Change Interventions Work Best in Combination.

Only a few health interventions with the general population have incorporated all three components of habit formation: response repetition, stable cues, and uncertain rewards. Yet, the few existing habit-based interventions that have bundled two or all three of these components have yielded promising results for weight loss[45] and consumption of healthy food in families.[46]

In one study, for example, overweight participants were instructed to (a) develop predictable and sustainable weight loss routines, (b) modify their home environments to increase cues to eat healthy foods and engage in exercise, and (c) have immediate positive rewards for weight-loss behaviors.[47] Participants also were instructed on how to disrupt existing habits by removing cues that triggered them along with making unhealthy behaviors less reinforcing (for example, increasing the preparation time and effort for unhealthy snacks). As depicted in Figure 2A, participants undergoing this multifaceted habit formation and disruption treatment continued to lose weight during several months following the end of the intervention, whereas participants using a more standard weight-loss program relapsed over time.

A very different habit formation intervention used an electronic monitoring device to promote weight loss among overweight adolescents.[48] This intervention targeted a specific behavior: the amount and speed of eating. Cues to eating were standardized by having participants undergo monitoring by a device while eating dinner at a table. The device delivered feedback about success and failure in hitting predetermined goals. As shown in Figure 2B, after 12 months,

Figure 2. Interventions specifically targeting habits can create enduring behavior change over time

In behavior change interventions that target habit formation and change, more enduring behavior change is possible.

A. Multifaceted habit formation and disruption weight loss program vs. standard weight loss program
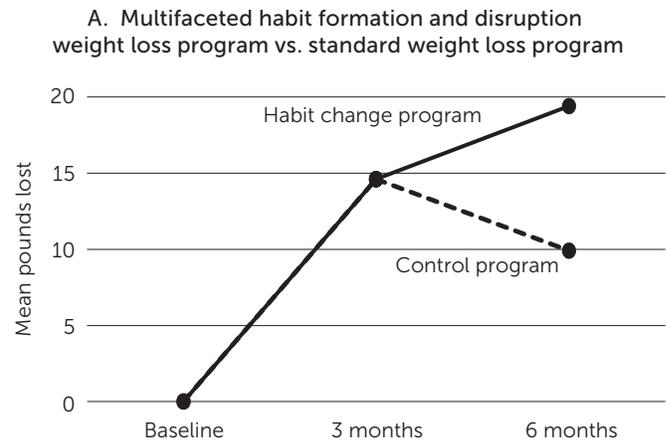
Figure A: Mean pounds lost after 3 months (mos) of habit-based or standard weight loss interventions (N = 59 at baseline, N = 35 at 6 months). The habit-based intervention emphasized (a) developing and maintaining healthy habits and disrupting unhealthy habits, (b) creating a personal food and exercise environment that increased exposure to healthy eating and physical activity and encouraged automatic responding to goal-related cues, and (c) facilitating weight loss motivation. The standard weight loss program involved examining attitudes toward food, body, and weight, such as improving body acceptance and understanding social stereotypes. Data are from "A Randomized Trial Comparing Two Approaches to Weight Loss: Differences in Weight Loss Maintenance," by R. A. Carels, J. M. Burmeister, A. M. Koball, M. W. Oehlhof, N. Hinman, M. LeRoy, . . . A. Gumble, 2014, *Journal of Health Psychology,* 19, p. 304, Figure 2. Copyright 2014 by Sage.

B. Electronic monitoring device to promote control of eating vs. standard weight loss program
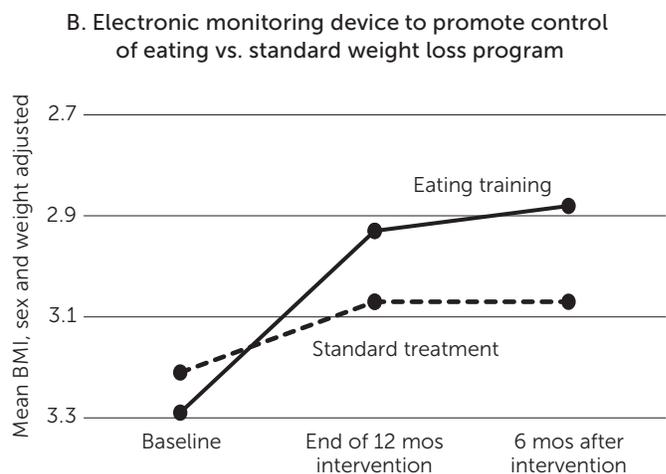
Figure B: Mean children's age- and sex-adjusted body mass index (BMI) after a yearlong intervention using a monitoring device to reduce the amount and speed of eating, plus a 6-month follow-up (N = 106 at baseline and 12 months, N = 87 at the 18-month assessment). Data are from "Treatment of Childhood Obesity by Retraining Eating Behaviour: Randomised Controlled Trial," by A. L. Ford, C. Bergh, P. Södersten, M. A. Sabin, S. Hollinghurst, L. P. Hunt, and J. P. Shield, 2010, *British Medical Journal,* 340, Article b5388, Table 2. Copyright 2010 by BMJ.

monitored participants not only ate smaller meals than participants in a control group did, but they had lost significant amounts of weight and kept it off 6 months after the intervention ended.

Breaking Unhealthy Habits

Because habits are represented in memory in a relatively separate manner from goals and conscious intentions, existing habits do not readily change when people adopt new goals. Thus, recognizing the health value of five servings of fruits and vegetables per day does not, by itself, remove the cues that trigger consumption of other less healthful foods. Similarly, incentive programs to break habits will not necessarily alter the memory trace underlying the behavior. Familiar contexts and routines still will bring unhealthy habits to mind, leaving people at risk of lapsing into old patterns.[49] Even after new habits have been formed, the existing memory traces are not necessarily replaced but instead remain dormant and can be reactivated relatively easily with a memory cue.[50]

Changing unhealthy habits, much like forming healthy ones, requires an understanding of the psychology behind habits. Specifically, ridding oneself of unhealthy habits requires neutralizing the context cues that automatically trigger habit performance.

The Three Main Habit-Breaking Interventions

Health interventions can incorporate three strategies to reduce the impact of existing bad cues: (a) cue disruption, (b) environmental reengineering, and (c) vigilant monitoring or inhibition (see Table 2). Experiments show that habit performance is readily disrupted when contexts have shifted.[50,51]

Cue Disruption

Interventions can take advantage of naturally occurring life events—such as moving to a new house, beginning a new job, or having a child—that reduce or eliminate exposure to the familiar cues that automatically trigger habit performance (see Table 2). People are most successful at changing their behavior in daily life when they capitalize on such life events. In a study in which people reported their attempts to change some unwanted behavior, moving to a new location

was mentioned in 36% of successful behavior change attempts but only in 13% of unsuccessful ones.[52] In addition, 13% of successful changers indicated that, to support the change, they altered the environment where a prior habit was performed, whereas none of the unsuccessful ones mentioned this.

Habit discontinuity interventions capitalize on this window of opportunity in which people are no longer exposed to cues that trigger old habits.[53] For example, an intervention that provided a free transit pass to car commuters increased the use of transit only among those who changed their residence or workplace in the prior 3 months.[54] Apparently, the move from a familiar environment disrupted cues to driving a car, enabling participants to act on the incentive to use transit instead of falling back on their car-driving habit. Another study showed that students' TV-watching habits were disrupted when they transferred to a new university, but only if cues specific to this behavior changed, such as their new residence no longer having a screen in the living room.[55] Without the old cue to trigger their TV habits, students only watched TV at the new university if they intended to.

Many different health interventions can be applied during the window of opportunity provided by life transitions. For example, new residents could be messaged, via text or mailers, with incentives to perform healthy behaviors related to their recent move. These could include reminders of the public transit options in the new neighborhood, notices that registration is open for community fitness classes, and invitations to local farmers' markets. Similarly, new employees could be informed about workplace-related health options such as employer-sponsored health classes. Also, reduced insurance rates could be offered if employees quit smoking or adopt other healthy behaviors. First-time parents could be engaged by interventions that encourage the preparation of healthy meals when cooking at home or that promote enrollment in child-and-parent exercise classes.

Environmental Reengineering

The impact of unhealthy habit cues also can be reduced by altering performance environments, or the place where the unhealthy habit regularly occurs (see Table 2). Although environmental reengineering often involves cue disruption (as described above), it additionally

---

Table 2. Three main components of habit-breaking interventions and examples of implementation in practice

Principle	Examples in practice
Cue disruption	• Target recent movers with public transit price reductions • Target new employees with health and wellness programs • Reduce salience of cues to unhealthy choices; increase salience of healthy choices (for example, redesign cafeterias to show healthy items first)
Environmental reengineering	Add friction to unhealthy behaviors • Banning smoking in public places • Banning visual reminders of cigarettes at point of purchase • Changing building design regulations to increase prominence of stairways • Explaining through public health communications how to alter personal environments to reduce the salience of unhealthy foods Remove friction from healthy behaviors • Starting bike-share programs • Bundling healthy food items in fast food menu selections (for example, apple slices as default side item) • Adding a fast check-out line in cafeterias for those purchasing healthy items only
Vigilant monitoring	• Food labeling regulations that require visual cues on packaging to show serving sizes • GPS technology triggers in smartphones and wearable devices that deliver nudges to adopt healthful behaviors (for example, based on time to and location of fast food restaurants, sending "don't go" alerts or "order this not that" messaging)

introduces new or altered environmental features to support the healthy behavior. The basic psychological process involves *adding behavioral friction* to unhealthy options and *reducing behavioral friction* for healthy ones to lubricate their adoption.

Adding friction. Large-scale social policies can introduce friction into an environment, making it harder for people to perform unhealthy habits. Smoking bans in English pubs, for instance, made it more difficult for people with strong smoking habits to light up while drinking.[56] Having to leave the pub to smoke creates friction, so smoking bans have generally increased quit rates.[57] Bans on visible retail displays of cigarettes also add friction by forcing potential purchasers to remember to request cigarettes.[58] Such bans are especially likely to reduce impulsive tobacco purchases[59] by removing environmental smoking cues.[60]

Another way of adding friction to unhealthy options is being tested in several cities in Switzerland. Policymakers are providing citizens with free electric bikes or free ride-share schemes, but only after they hand over their car keys for a few weeks. The idea is to add friction to existing car-use habits.[61] If successful, blocking the automatic response of car driving will encourage the use of other forms of transit that, in turn, may become habitual.

Reducing friction. A variety of existing policies successfully alter physical environments to promote frictionless accessibility to healthy behaviors over unhealthy ones. These include the availability of recreational facilities, opportunities to walk and cycle, and accessibility of stores selling fresh foods. The effectiveness of such friction-easing interventions is clear: U.S. residents with access to parks closer to home engage in more leisure-time physical activity and have lower rates of obesity.[62] Also, a bike-share program instituted in London increased exercise rates.[63] Furthermore, in U.S. metropolitan areas, fruit and vegetable consumption was greater and obesity rates were lower among people living closer to a supermarket with fresh foods.[64]

The broad success of environmental reengineering policies and changes to the physical environment makes these prime strategies for large-scale habit change. Nonetheless, these initiatives require political and citizen support for healthy policies, tax codes, and zoning. We suspect that such support will increase in the future, given increasing recognition of lifestyle effects on health.[65] To illustrate this potential, we note that building

codes could make healthy options the default choice by applying friction to elevator use so that stairways are readily accessible and elevators less apparent. In addition, to add friction to unhealthy food choices and to automate healthy ones, restaurants could provide food bundles (for example, value meals) with healthy default options (for example, apple slices instead of French fries), and manufacturers could switch to packaging formats that do not minimize apparent food quantity but enable people to accurately assess the amount they are eating.[66] To simplify consumer understanding of healthy choices, restaurants and food companies could be rated for health performance, much as they currently are for sanitation.[67]

Finally, on a more immediate, personal level, behavior change interventions can provide individuals with the knowledge and ability to reengineer their own personal environments. The potential benefits of change in microenvironments have been demonstrated clearly with respect to healthy eating: People with a lower body mass index were likely to have fruit available on their kitchen counters, whereas those weighing more were likely to have candy, sugary cereal, and nondiet soft drinks.[68] And demonstrating that food choice is based in part on high visibility, studies that have directly manipulated the visibility and convenience of foods reveal that people tend to consume easily accessible, frictionless options rather than inaccessible, high-friction choices.[69] Another approach to reduce the friction to healthy choices is allowing people to preorder food, enabling them to make healthier choices outside of the influence of the evocative smells and visual temptations of school or work cafeterias.[70] In summary, it is sound policy to empower individuals to reengineer their immediate environments to increase access to contexts promoting healthy behaviors and avoid contexts of unhealthy ones.

Vigilant Monitoring

Inhibition of habits through *vigilant monitoring* is a final habit-breaking strategy that increases awareness of the cues that trigger unhealthy habits and provides opportunities to inhibit them (see Table 2). Unlike cue disruption and environmental reengineering, which focus primarily on harnessing automatic processes, vigilant monitoring combines conscious thoughts of control with automatic processes. This works as a sort of cognitive override process.

Vigilant monitoring is the strategy that people are most likely to use to control unwanted habits in daily life.[71] By thinking, "Don't do it," and monitoring carefully for slipups, participants in several studies were more effective at curbing bad habits such as eating junk food, smoking, and drinking too much than when they used other strategies (for example, distracting themselves). These researchers subsequently brought this strategy into the lab to study it under controlled conditions using a word-pair task. Vigilant monitoring proved to control habits by heightening inhibitory cognitive control processes at critical times when bad habits were most likely—that is, by helping people combat their automatic responses before they happened.

Vigilance may be most effective when paired with strategies that also make healthy options cognitively accessible, so the desired action is salient in contexts in which people have an unhealthy habit. Thus, after people formed implementation intentions to eat apples or another healthy snack in a context in which they typically ate unhealthy ones like candy bars, the healthy behavior automatically came to mind when that context was encountered in the future.[23]

Facilitating vigilant monitoring for individuals. Because vigilant inhibition is effortful to sustain, it could be facilitated by GPS technology in smartphones and wearable devices that enable reminders or *nudges*, to be delivered on the basis of physical proximity to locations linked with unwanted habits (for example, fast food restaurants). Given that these sensor devices can detect daily activities such as eating and watching TV,[72] they could potentially deliver response-timed electronic prompts at just the right time to inhibit acting on unhealthy habits.

In policy applications, vigilant monitoring of unwanted behaviors can be adapted into interventions through reminders to control unwanted habits. These could be conveyed indirectly with simple changes to product packaging, such as pictures illustrating the amount of a single-serving portion on a bag of Oreos. Or serving cues could be embedded within the food itself, perhaps by inserting a different-colored cookie at a certain point in the package to trigger a "stop here" response.[73] More directly, *point-of-choice prompts* involving signs or other reminders of desired actions might be used in situations where people usually respond in other ways. For example, signs to promote stair climbing over elevator and escalator use in public

settings have shown modest but consistent success.[74] Because such reminders may become less effective over time, except among people who perform the behavior sufficiently often so that it becomes habitual,[75] it may be necessary to diversify such visual cues over time to help retrigger vigilance.

Framework for Policymakers

Habit-based interventions are tailored to the mechanisms of action, ensuring that the patterning of behavior is optimal to create healthy habits and impede unhealthy ones. The principles and tactics outlined here can be applied at varying levels of scale, with some best suited to individual self-change, others to community health interventions, and still others to state and national policies. So, which of the ideas we have discussed in this article scale best for public policy?

For Habit Formation

Public policy regulations can effectively make healthy responses salient (for example, funding bike paths and bike-share programs) and tie desired behaviors to stable contexts (for example, public health communications that link reminders to change smoke detector batteries to the start and end of daylight savings time, medical compliance communications that piggyback medication intake onto an existing habit). At its core, habit formation is promoted through the various public policies that incentivize repeated healthy responses in stable contexts (for example, free public transit days; Supplemental Nutrition Assistance Program benefits limited to the purchase of high nutrition, low-energy-dense foods such as spinach and carrots).

For Habit Disruption

Policymakers can initiate legislation to reduce the presence of unhealthy habit cues (for example, funding the reengineering of school cafeterias) and can also harness context disruption (for example, free public transit programs for recent movers). The success of anti-smoking campaigns provides a model for how this can work. Among the many different policies used to control tobacco, the most successful were the ones that added friction to smoking, such as increasing tobacco prices, instituting smoking bans in public places, and removing

tobacco and advertising from point-of-purchase displays.[9] As would be anticipated given the habitual, addictive nature of smoking, warning labels on packets have limited impact,[65] and mass media campaigns have generally only been effective in conjunction with the more friction-inducing interventions listed above.[76,77]

Traditional policy tools such as tax breaks are a generally useful tool for health behavior change. Linking tax breaks for health insurers to policyholders' health habits can create incentives for companies and other large institutions to apply habit-change principles in more localized ways. Tax policies can also drive habit change by adding friction to unhealthy consumer choices (for example, taxes on sugared soft drinks, tobacco, and fast food).

For many everyday health challenges, people are likely to benefit from both forming healthy habits and disrupting unhealthy ones. Thus, multicomponent interventions that include distinct elements designed to break existing habits and support the initiation and maintenance of new ones will be needed. For example, an intervention to increase fruit and vegetable consumption among students in a school cafeteria could simultaneously reengineer the choice environment to disrupt their existing habits to eat processed snacks (for example, by moving such snacks to the back of displays and fruit to the front) and to form new habits (for example, by providing discounts to incentivize the selection and consumption of healthful foods, or express checkout lanes for people making healthy purchases). However, habit disruption is, of course, irrelevant in shifting, changing environments and for people who do not have a history of acting in a given domain or circumstance. Thus, habit interruptions have more limited use than the broadly applicable habit formation principles.

Conclusion

Strategies that accelerate habit formation and promote maintenance are especially important for health interventions, given that many benefits of healthy behaviors are not evident immediately but instead accrue gradually with repetition. Thus, interventions that are successful at promoting short spurts of exercise or a sporadically healthful diet will provide little protection against the risks of lifestyle diseases associated with inactivity and overeating. The habit-based strategies outlined in

this article provide policymakers and behavior change specialists with important insights into the mechanisms by which people can create sustainable healthy lifestyles.

author affiliation

Wood, Dornsife Department of Psychology and Marshall School of Business, University of Southern California; Neal, Catalyst Behavioral Sciences and Center for Advanced Hindsight, Duke University. Corresponding author's e-mail: wendy.wood@usc.edu

References

1. Stables, G. J., Subar, A. F., Patterson, B. H., Dodd, K., Heimendinger, J., Van Duyn, M. A. S., & Nebeling, L. (2002). Changes in vegetable and fruit consumption and awareness among US adults: Results of the 1991 and 1997 5 A Day for Better Health Program surveys. *Journal of the American Dietetic Association, 102,* 809–817. http://dx.doi.org/10.1016/S0002-8223(02)90181-1

2. Casagrande, S. S., Wang, Y., Anderson, C., & Gary, T. L. (2007). Have Americans increased their fruit and vegetable intake? The trends between 1988 and 2002. *American Journal of Preventive Medicine, 32,* 257–263. http://dx.doi.org/10.1016/j.amepre.2006.12.002

3. Moore, L. V., & Thompson, F. E. (2015, July 10). Adults meeting fruit and vegetable intake recommendations—United States, 2013. *Morbity and Mortality Weekly Report, 64,* 709–713. http://www.cdc.gov/MMWR/preview/mmwrhtml/mm6426a1.htm

4. Vandelanotte, C., Spathonis, K. M., Eakin, E. G., & Owen, N. (2007). Website-delivered physical activity interventions: A review of the literature. *American Journal of Preventive Medicine, 33,* 54–64. http://dx.doi.org/10.1016/j.amepre.2007.02.041

5. Fjeldsoe, B., Neuhaus, M., Winkler, E., & Eakin, E. (2011). Systematic review of maintenance of behavior change following physical activity and dietary interventions. *Health Psychology, 30,* 99–109. http://dx.doi.org/10.1037/a0021974

6. Mantzari, E., Vogt, F., Shemilt, I., Wei, Y., Higgins, J. P., & Marteau, T. M. (2015). Personal financial incentives for changing habitual health-related behaviors: A systematic review and meta-analysis. Preventive Medicine, 75, 75-85. http://dx.doi.org/10.1016/j.ypmed.2015.03.001

7. Volpp, K. G., John, L. K., Troxel, A. B., Norton, L., Fassbender, J., & Loewenstein, G. (2008). Financial incentive−based approaches for weight loss: A randomized trial. *Journal of the American Medical Association, 300,* 2631–2637. http://dx.doi.org/10.1001/jama.2008.804

8. Charness, G., & Gneezy, U. (2009). Incentives to exercise. *Econometrica, 77,* 909–931. http://dx.doi.org/10.3982/ECTA7416

9. Volpp, K. G., Troxel, A. B., Pauly, M. V., Glick, H. A., Puig, A., Asch, D. A., . . . Audrain-McGovern, J. (2009). A randomized, controlled trial of financial incentives for smoking cessation. *New England Journal of Medicine, 360,* 699–709. http://dx.doi.org/10.1056/NEJMsa0806819

10. King, A. C., Hekler, E. B., Castro, C. M., Buman, M. P., Marcus, B. H., Friedman, R. H., & Napolitano, M. A. (2014). Exercise advice by humans versus computers: Maintenance effects at 18 months. *Health Psychology, 33,* 192–196. http://dx.doi.org/10.1037/a0030646

11. Adriaanse, M. A., Kroese, F. M., Gillebaart, M., & De Ridder, D. T. (2014). Effortless inhibition: Habit mediates the relation between self-control and unhealthy snack consumption. *Frontiers in Psychology, 5,* Article 444. http://dx.doi.org/10.3389/fpsyg.2014.00444

12. Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology, 109,* 508–525. http://dx.doi.org/10.1037/pspp0000026

13. Wilson, L. M., Tang, E. A., Chander, G., Hutton, H. E., Odelola, O. A., Elf, J. L., . . . Apelberg, B. J. (2012). Impact of tobacco control interventions on smoking initiation, cessation, and prevalence: A systematic review. *Journal of Environmental and Public Health, 2012,* Article 961724. http://dx.doi.org/10.1155/2012/961724

14. Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29,* 2225–2232. http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x

15. Neal, D. T., Wood, W., Wu, M., & Kurlander, D. (2011). The pull of the past: When do habits persist despite conflict with motives? *Personality and Social Psychology Bulletin, 37,* 1428–1437. http://dx.doi.org/10.1177/0146167211419863

16. Lin, P.-Y., Wood, W., & Monterosso, J. (2016). Healthy eating habits protect against temptations. *Appetite, 103,* 432–440. http://dx.doi.org/10.1016/j.appet.2015.11.011

17. Neal, D. T., Wood, W., Labrecque, J. S., & Lally, P. (2012). How do habits guide behavior? Perceived and actual triggers of habits in daily life. *Journal of Experimental Social Psychology, 48,* 492–498. http://dx.doi.org/10.1016/j.jesp.2011.10.011

18. Thaler, R. H., Sunstein, C. R., & Balz, J. P. (2012). Choice architecture. In E. Shafir (Ed.), *The behavioral foundations of public policy* (pp. 428–439). Princeton, NJ: Princeton University Press. http://dx.doi.org/10.2139/ssrn.2536504

19. Johnson, E. J., Shu, S. B., Dellaert, B. G. C., Fox, C., Goldstein, D. G., Häubl, G., . . . Weber, E. U. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters, 23,* 487–504. http://dx.doi.org/10.1007/s11002-012-9186-1

20. Sherif, M. (1936). *The psychology of social norms.* Oxford, England: Harper.

21. Salmon, S. J., Fennis, B. M., de Ridder, D. T., Adriaanse, M. A., & de Vet, E. (2014). Health on impulse: When low self-control promotes healthy food choices. *Health Psychology, 33,* 103–109. http://dx.doi.org/10.1037/a0031785

22. Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist, 54,* 493–503. http://dx.doi.org/10.1037/0003-066X.54.7.493

23. Adriaanse, M. A., Gollwitzer, P. M., de Ridder, D. T. D., de Wit, J. B. F., & Kroese, F. M. (2011). Breaking habits with implementation intentions: A test of underlying processes. *Personality and Social Psychology Bulletin, 37,* 502–513. http://dx.doi.org/10.1177/0146167211399102

24. Wood, W., & Rünger, D. (2016). The psychology of habit. *Annual Review of Psychology, 67,* 289–314. http://dx.doi.org/10.1146/annurev-psych-122414-033417

25. Ji, M. F., & Wood, W. (2007). Purchase and consumption habits: Not necessarily what you intend. *Journal of Consumer Psychology, 17,* 261–276. http://dx.doi.org/10.1016/S1057-7408(07)70037-2

26. Verhoeven, A. A. C., Adriaanse, M. A., Evers, C., & de Ridder, D. T. D. (2012). The power of habits: Unhealthy snacking behaviour is primarily predicted by habit strength. *British Journal of Health Psychology, 17,* 758–770. http://dx.doi.org/10.1111/j.2044-8287.2012.02070.x

27. Neal, D. T., Wood, W., & Drolet, A. (2013). How do people adhere to goals when willpower is low? The profits (and pitfalls) of strong habits. *Journal of Personality and Social Psychology, 104,* 959–975. http://dx.doi.org/10.1037/a0032626

28. Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology, 40,* 998–1009. http://dx.doi.org/10.1002/ejsp.674

29. Armitage, C. J. (2005). Can the theory of planned behavior predict the maintenance of physical activity? *Health Psychology, 24,* 235–245. http://dx.doi.org/10.1037/0278-6133.24.3.235

30. Kaushal, N., & Rhodes, R. E. (2015). Exercise habit formation in new gym members: A longitudinal study. *Journal of Behavioral Medicine, 38,* 652–663. http://dx.doi.org/10.1007/s10865-015-9640-7

31. Neal, D. T., Vujcic, J., Hernandez, O., & Wood, W. (2015). *Creating hand-washing habits: Six principles for creating disruptive and sticky behavior change for hand washing with soap.* Unpublished manuscript, Catalyst Behavioral Science, Miami, FL.

32. Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review, 104,* 3003–3037. http://dx.doi.org/10.1257/aer.104.10.3003

33. Tappe, K., Tarves, E., Oltarzewski, J., & Frum, D. (2013). Habit formation among regular exercisers at fitness centers: An exploratory study. *Journal of Physical Activity & Health, 10,* 607–613.

34. Brooks, T. L., Leventhal, H., Wolf, M. S., O'Conor, R., Morillo, J., Martynenko, M., Wisnivesky, J. P., & Federman, A. D. (2014). Strategies used by older adults with asthma for adherence to inhaled corticosteroids. *Journal of General Internal Medicine, 29,* 1506–1512. http://dx.doi.org/10.1007/s11606-014-2940-8

35. Rogers, T., Milkman, K. L., John, L. K., & Norton, M. I. (2015). Beyond good intentions: Prompting people to make plans improves follow-through on important tasks. *Behavioral Science & Policy, 1*(2), 33–41.

36. Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology, 29,* 374–383. http://dx.doi.org/10.1037/a0019596

37. Judah, G., Gardner, B., & Aunger, R. (2013). Forming a flossing habit: An exploratory study of the psychological determinants of habit formation. *British Journal of Health Psychology, 18,* 338–353. http://dx.doi.org/10.1111/j.2044-8287.2012.02086.x

38. Labrecque, J. S., Wood, W., Neal, D. T., & Harrington, N. (2016). Habit slips: When consumers unintentionally resist new products. *Journal of the Academy of Marketing Science.* Advance online publication. http://dx.doi.org/10.1007/s11747-016-0482-9

39. Phillips, A. L., Leventhal, H., & Leventhal, E. A. (2013). Assessing theoretical predictors of long-term medication adherence: Patients' treatment-related beliefs, experiential feedback and habit development. *Psychology & Health, 28,* 1135–1151. http://dx.doi.org/10.1080/08870446.2013.793798

40. Lally, P., & Gardner, B. (2013). Promoting habit formation. *Health Psychology Review, 7*(Suppl. 1), S137–S158. http://dx.doi.org/10.1080/17437199.2011.603640

41. Fishbach, A., & Trope, Y. (2005). The substitutability of external control and self-control. *Journal of Experimental Social Psychology, 41,* 256–270. http://dx.doi.org/10.1016/j.jesp.2004.07.002

42. DeRusso, A. L., Fan, D., Gupta, J., Shelest, O., Costa, R. M., & Yin, H. H. (2010). Instrumental uncertainty as a determinant of behavior under interval schedules of reinforcement. *Frontiers in Integrative Neuroscience, 4,* Article 17. http://dx.doi.org/10.3389/fnint.2010.00017

43. Burns, R. J., Donovan, A. S., Ackermann, R. T., Finch, E. A., Rothman, A. J., & Jeffery, R. W. (2012). A theoretically grounded systematic review of material incentives for weight loss: Implications for interventions. *Annals of Behavioral Medicine, 44,* 375–388. http://dx.doi.org/10.1007/s12160-012-9403-4

44. Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *The Journal of Economic Perspectives, 25,* 191–209. http://dx.doi.org/10.1257/jep.25.4.191

45. Lally, P., Chipperfield, A., & Wardle, J. (2008). Healthy habits: Efficacy of simple advice on weight control based on a habit-formation model. *International Journal of Obesity, 32,* 700–707. http://dx.doi.org/10.1038/sj.ijo.0803771

46. Gardner, B., Sheals, K., Wardle, J., & McGowan, L. (2014). Putting habit into practice, and practice into habit: A process evaluation and exploration of the acceptability of a habit-based dietary behaviour change intervention. *International Journal of Behavioral Nutrition and Physical Activity, 11,* Article 135. http://dx.doi.org/10.1186/s12966-014-0135-7

47. Carels, R. A., Burmeister, J. M., Koball, A. M., Oehlhof, M. W., Hinman, N., LeRoy, M., . . . Gumble, A. (2014). A randomized trial comparing two approaches to weight loss: Differences in weight loss maintenance. *Journal of Health Psychology, 19,* 296–311. http://dx.doi.org/10.1177/1359105312470156

48. Ford, A. L., Bergh, C., Södersten, P., Sabin, M. A., Hollinghurst, S., Hunt, L. P., & Shield, J. P. (2010). Treatment of childhood obesity by retraining eating behaviour: Randomised controlled trial. *British Medical Journal, 340,* Article b5388. http://dx.doi.org/10.1136/bmj.b5388

49. Walker, I., Thomas, G. O., & Verplanken, B. (2015). Old habits die hard: Travel habit formation and decay during an office relocation. *Environment Behavior, 47,* 1089–1106. http://dx.doi.org/10.1177/0013916514549619

50. Bouton, M. E., Todd, T. P., Vurbic, D., & Winterbauer, N. E. (2011). Renewal after the extinction of free operant behavior. *Learning & Behavior, 39,* 57–67. http://dx.doi.org/10.3758/s13420-011-0018-6

51. Thrailkill, E. A., & Bouton, M. E. (2015). Extinction of chained instrumental behaviors: Effects of procurement extinction on consumption responding. *Journal of Experimental Psychology: Animal Learning and Cognition, 41,* 232–246. http://dx.doi.org/10.1037/xan0000064

52. Heatherton, T. F., & Nichols, P. A. (1994). Personal accounts of successful versus failed attempts at life change. *Personality and Social Psychology Bulletin, 20,* 664–675. http://dx.doi.org/10.1177/0146167294206005

53. Verplanken, B., Walker, I., Davis, A., & Jurasek, M. (2008). Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses. *Journal of Environmental Psychology, 28,* 121–127. http://dx.doi.org/10.1016/j.jenvp.2007.10.005

54. Thøgersen, J. (2012). The importance of timing for breaking commuters' car driving habits. *Collegium: Studies Across*

Disciplines in the Humanities and Social Sciences, 12, 130–140. Retrieved from https://helda.helsinki.fi/bitstream/handle/10138/34227/12_08_thogersen.pdf?sequence=1

55. Wood, W., Tam, L., & Witt, M. G. (2005). Changing circumstances, disrupting habits. *Journal of Personality and Social Psychology, 88,* 918–933. http://dx.doi.org/10.1037/0022-3514.88.6.918

56. Orbell, S., & Verplanken, B. (2010). The automatic component of habit in health behavior: Habit as cue-contingent automaticity. *Health Psychology, 29,* 374–383. http://dx.doi.org/10.1037/a0019596

57. Lemmens, V., Oenema, A., Knut, I. K., & Brug, J. (2008). Effectiveness of smoking cessation interventions among adults: A systematic review of reviews. *European Journal of Cancer Prevention, 17,* 535–544. http://dx.doi.org/10.1097/CEJ.0b013e3282f75e48

58. Wakefield, M., Germain, D., & Henriksen, L. (2008). The effect of retail cigarette pack displays on impulse purchase. *Addiction, 103,* 322–328. http://dx.doi.org/10.1111/j.1360-0443.2007.02062.x

59. Robertson, L., McGee, R., Marsh, L., & Hoek, J. (2014). A systematic review on the impact of point-of-sale tobacco promotion on smoking. *Nicotine & Tobacco Research, 17,* 2–17. http://dx.doi.org/10.1093/ntr/ntu168

60. Kirchner, T. R., Cantrell, J., Anesetti-Rothermel, A., Ganz, O., Vallone, D. M., & Abrams, D. B. (2013). Geospatial exposure to point-of-sale tobacco: Real-time craving and smoking-cessation outcomes. *American Journal of Preventive Medicine, 45,* 379–385. http://dx.doi.org/10.1016/j.amepre.2013.05.016

61. Lourenço, J. S., Ciriolo, E., Almeida, S. R., & Troussard, X. (2016). *Behavioural insights applied to policy: European Report 2016* (Report No. EUR 27726 EN). http://dx.doi.org/10.2760/903938

62. Roubal, A. M., Jovaag, A., Park, H., & Gennuso, K. P. (2015). Development of a nationally representative built environment measure of access to exercise opportunities. *Preventing Chronic Disease, 12,* Article 140378. http://dx.doi.org/10.5888/pcd12.140378

63. Woodcock, J., Tainio, M., Cheshire, J., O'Brien, O., & Goodman, A. (2014). Health effects of the London bicycle sharing system: Health impact modelling study. *British Medical Journal, 348,* Article g425. http://dx.doi.org/10.1136/bmj.g425

64. Michimi, A., & Wimberly, M. C. (2010). Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. *International Journal of Health Geographics, 9,* Article 49. http://dx.doi.org/10.1186/1476-072X-9-49

65. Kohl, H. W., Craig, C. L., Lambert, E. V., Inoue, S., Alkandari, J. R., Leetongin, G., . . . Lancet Physical Activity Series Working Group. (2012). The pandemic of physical inactivity: Global action for public health. *The Lancet, 380,* 294–305. http://dx.doi.org/10.1016/S0140-6736(12)60898-8

66. Krishna, A. (2006). Interaction of senses: The effect of vision versus touch on the elongation bias. *Journal of Consumer Research, 32,* 557–566.

67. Cohen, D., Bhatia, R., Story, M. T., Wootan, M., Economos, C. D., Van Horn, L., . . . Williams, J. D. (2013). *Performance standards for restaurants: A new approach to addressing the obesity epidemic*. Retrieved from http://www.rand.org/pubs/conf_proceedings/CF313.html

68. Wansink, B., Hanks, A. S., & Kaipainen, K. (2015). Slim by design: Kitchen counter correlates of obesity. *Health Education & Behavior.* Advance online publication. http://dx.doi.org/10.1177/1090198115610571

69. Rozin, P., Scott, S., Dingley, M., Urbanek, J. K., Jiang, H., & Kaltenbach, M. (2011). Nudge to nobesity I: Minor changes in accessibility decrease food intake. *Judgment and Decision Making, 6,* 323–332.

70. Hanks, A. S., Just, D. R., & Wansink, B. (2013). Preordering school lunch encourages better food choices by children. *JAMA Pediatrics, 167,* 673–674. http://dx.doi.org/10.1001/jamapediatrics.2013.82

71. Quinn, J. M., Pascoe, A., Wood, W., & Neal, D. T. (2010). Can't control yourself? Monitor those bad habits. *Personality and Social Psychology Bulletin, 36,* 499–511. http://dx.doi.org/10.1177/0146167209360665

72. Chen, G., Ding, X., Huang, K., Ye, X., & Zhang, C. (2015, February). *Changing health behaviors through social and physical context awareness*. Paper presented at the International Conference on Computing, Networking, and Communications, Anaheim, CA.

73. Geier, A., Wansink, B., & Rozin, P. (2012). Red potato chips: Segmentation cues substantially decrease food intake. *Health Psychology, 31,* 398–401.

74. Soler, R. E., Leeks, K. D., Buchanan, L. R., Brownson, R. C., Heath, G. W., Hopkins, D. H., & Task Force on Community Preventive Services. (2010). Point-of-decision prompts to increase stair use. *American Journal of Preventive Medicine, 38*(2, Suppl.), S292–S300. http://dx.doi.org/10.1016/j.amepre.2009.10.028

75. Tobias, R. (2009). Changing behavior by memory aids: A social psychological model of prospective memory and habit development tested with dynamic field data. *Psychological Review, 116,* 408–438. http://dx.doi.org/10.1037/a0015512

76. Bala, M., Strzeszynski, L., & Cahill, K. (2008). Mass media interventions for smoking cessation in adults. *Cochrane Database of Systematic Reviews, 2013*(6), Article CD004704. http://dx.doi.org/10.1002/14651858.CD004704.pub3

77. Levy, D. T., Chaloupka, F., & Gitchell, J. (2004). The effects of tobacco control policies on smoking rates: A tobacco control scorecard. *Journal of Public Health Management and Practice, 10,* 338–353.

Making the truth stick & the myths fade: Lessons from cognitive psychology

Norbert Schwarz, Eryn Newman, & William Leach

Summary. Erroneous beliefs are difficult to correct. Worse, popular correction strategies, such as the myth-versus-fact article format, may backfire because they subtly reinforce the myths through repetition and further increase the spread and acceptance of misinformation. Here we identify five key criteria people employ as they evaluate the truth of a statement: They assess general acceptance by others, gauge the amount of supporting evidence, determine its compatibility with their beliefs, assess the general coherence of the statement, and judge the credibility of the source of the information. In assessing these five criteria, people can actively seek additional information (an effortful analytic strategy) or attend to the subjective experience of easy mental processing—what psychologists call *fluent processing*—and simply draw conclusions on the basis of what feels right (a less effortful intuitive strategy). Throughout this truth-evaluation effort, fluent processing can facilitate acceptance of the statement: When thoughts flow smoothly, people nod along. Unfortunately, many correction strategies inadvertently make the false information more easily acceptable by, for example, repeating it or illustrating it with anecdotes and pictures. This, ironically, increases the likelihood that the false information the communicator wanted to debunk will be believed later. A more promising correction strategy is to focus on making the true information as easy to process as possible. We review recent research and offer recommendations for more effective presentation and correction strategies.

Back in 2000, flesh-eating bananas were on the loose and wreaking havoc, according to trending Internet reports. The story claimed that exported bananas contained necrotizing bacteria that could infect consumers after they had eaten the fruit. It was a hoax, but one with such legs of believability that the Centers for Disease Control and Prevention (CDC) set up a hotline to counter the misinformation and assure concerned fruit lovers that bananas were perfectly safe. The *Los Angeles Times* even ran an article explaining the

origin of the myth, noting that the hoax gained traction because a secretary from the University of California, Riverside's agricultural college forwarded the story to friends in an e-mail, seemingly giving it the imprimatur of the college. Paradoxically, the efforts by the CDC and the *Los Angeles Times* to dispel the myth actually increased some people's acceptance of it, presumably because these trustworthy sources had taken the time and effort to address the "problem." These corrections likely made the myth more familiar and probably helped the myth and its variants to persist for the entire decade.[1]

No one doubts that the Internet can spread misinformation, but when such falsehoods go beyond banana hoaxes and into the health care realm, they have the potential to do serious harm. For example, websites abound that mischaracterize the scientific evidence and misstate the safety of vaccines, such as that they cause infection that can be passed on;[2] that falsely claim a certain kind of diet can beat back cancer, such as claims that drinking red wine can prevent breast cancer;[3] and that overstate preliminary associations between certain foods and healthful outcomes, such as that eating grapefruit burns fat.[4] These erroneous statements can cause people to modify their behaviors—perhaps in a detrimental fashion—affecting what they eat and how they seek medical care.

The persistence of the necrotizing banana myth shows that correcting false beliefs is difficult and that correction attempts often fail because addressing misinformation actually gives it more airtime, increasing its familiarity and making it seem even more believable.[5] For instance, one of the most frequently used correction strategies, the myth-versus-fact format, can backfire because of repetition of the myth, leaving people all the more convinced that their erroneous beliefs are correct.[6] The simple repetition of a falsehood, even by a questionable source, can lead people to actually believe the lie. The psychological research showing how people determine whether something is likely to be true has important implications for health communication strategies and can help point to more efficient approaches to disseminating well-established truths in general. Overall, behavioral research shows that often the best strategy in the fight against misinformation is to paint a vivid and easily understood summation of the truthful message one wishes to impart instead of drawing further attention to false information.

The Big Five Questions We Ask to Evaluate Truth

When people encounter a claim, they tend to evaluate its truth by focusing on a limited number of criteria.[7] Most of the time, they ask themselves at least one of five questions (see Table 1).

1. Social Consensus: Do Others Believe It?

In 1954, the American social psychologist Leon Festinger theorized that when the truth is unclear, people often turn to social consensus as a gauge for what is likely to be correct.[8] After all, if many people believe a claim, then there is probably something to it. A fun example of this is played out on the popular TV show *Who Wants to Be a Millionaire?* where, when stumped for the correct answer to a question, the contestant may poll the audience to see if there is a consensus answer.

Overall, people are more confident in their beliefs if others share them,[9,10] trust their memories more if others remember an event the same way,[11,12] and are more inclined to believe scientific theories if a consensus among scientists exists.[13]

To verify a statement's social consensus, people may turn to opinion polls, databases, or other external resources. Alternatively, they may simply ask themselves how often they have heard this belief. Chances are that a person is more frequently exposed to widely shared beliefs than to beliefs that are held by few others, so frequency of exposure should be a good gauge for a belief's popularity. Unfortunately, people are bad at tracking how often they have heard something and from whom; instead, people rely on whether a message feels familiar. This reliance gives small but vocal groups a great advantage: The more often they repeat their message, the more familiar it feels, leaving the impression that many people share the opinion.

For example, Kimberlee Weaver of Virginia Polytechnic Institute and her colleagues showed study participants a group discussion regarding public space.[14] The discussion presented the opinion that open spaces are desirable because they provide the community with opportunities for outdoor recreation. Participants heard the opinion either once or thrice, with a crucial difference: In one condition, three different people offered the opinion, whereas in the other condition, the same person repeated the opinion three times. Not surprisingly, participants thought that the opinion had broader

Table 1. Five criteria people use for judging truth

Criteria	Analytic evaluation	Intuitive evaluation
Social consensus: Do others believe it?	Search databases, look for supporting statistics, or poll a group or audience.	Does it feel familiar?
Support: Is there much supporting evidence?	Look for corroborating evidence in peer-reviewed scientific articles or news reports, or use one's own memory.	Is the evidence easy to generate or recall?
Consistency: Is it compatible with what I believe?	Recall one's own general knowledge and assess the match or mismatch with new information.	Does it make me stumble? Is it difficult to process, or does it feel right?
Coherence: Does it tell a good story?	Do the elements of the story logically fit together?	Does the story flow smoothly?
Credibility: Does it come from a credible source?	Is the source an expert? Does the source have a competing interest?	Does this source seem familiar and trustworthy?

support when three speakers offered it than when only one speaker did. But hearing the same statement three times from the same person was almost as influential as hearing it from three separate speakers, proving that a single repetitive voice can sound like a chorus.[14,15] These findings also suggest that the frequent repetition of the same sound bite in TV news or ads may give the message a familiarity that makes viewers overestimate its popularity. This is also the case on social media, where the same message keeps showing up as friends and friends of friends like it and repost it, resulting in many exposures within a network.

2. Support: Is There Much Evidence to Substantiate It?

When a large body of evidence supports a position, people are likely to trust it and believe that it is true. They can find this evidence through a deliberate search by looking for evidence in peer-reviewed scientific articles, reading substantiated news reports, or even combing their own memories. But people can also take a less taxing, speedier approach by making a judgment on the basis of how easy it is to retrieve or obtain some pieces of evidence. After all, the more evidence exists, the easier it should be to think of some. Indeed, when recalling evidence feels difficult, people conclude that there is less of it, regardless of how much information they actually remember. In one 1993 study,[16] Fritz Strack and Sabine Stepper, then of the University of Mannheim in Germany, asked participants to recall five instances in

which they behaved very assertively. To induce a feeling of difficulty, some were asked to furrow their eyebrows, an expression often associated with difficult tasks. When later asked how assertive they are, those who had to furrow their eyebrows judged themselves to be less assertive than did those who did not have to furrow their brows. Even though both groups recalled five examples of their own assertive behavior, they arrived at different conclusions when recall felt difficult.

In fact, the feeling of difficulty can even override the implications of coming up with a larger number of examples. In another study,[17] participants recalled just a few or many examples of their own assertive behavior. Whereas participants reported that recalling a few examples was easy, they reported that recalling many examples was difficult. As a result, those who remembered more examples of their own assertiveness subsequently judged themselves to be less assertive than did those who had to recall only a few examples. The difficulty of bringing many examples to mind undermined the examples' influence.

These findings have important implications for correction strategies. From a rational perspective, thinking of many examples or arguments should be more persuasive than thinking of only a few. Hence, correction strategies often encourage people to think of reasons why an erroneous or potentially erroneous belief may not hold.[18] But the more people try to do so, the harder it feels, leaving them all the more convinced that their belief is correct.[6] For example, in

a study described in an article published in the *Journal of Experimental Psychology; Learning , Memory, and Cognition*, participants read a short description of a historic battle in Nepal.[19] Some read that the British army won the battle, and others read that the Nepal Gurkhas won the battle. Next, they had to think about how the battle could have resulted in a different outcome. Some had to list only two reasons for a different outcome, whereas others had to list 10. Although participants in the latter group came up with many more reasons than did those in the former group for why the battle could have had a different result, they nevertheless thought that an alternative outcome was less likely. Such findings illustrate why people are unlikely to believe evidence that they find difficult to retrieve or generate: A couple of arguments that readily pop into the head are more compelling than many arguments that were hard to think of. As a result, simple and memorable claims have an advantage over considerations of a more complicated notion or reality.

3. Consistency: Is It Compatible with What I Believe?

People are inclined to believe things that are consistent with their own beliefs and knowledge.[20–22] One obvious way to assess belief consistency would be to recall general knowledge and assess its match with new information. For example, if you heard someone claim that vaccinations cause autism, you may check that claim against what you already know about vaccinations. But again, reliance on one's feelings while thinking about the new information provides an easier route to assessing consistency. When something is inconsistent with existing beliefs, people tend to stumble—they take longer to read it and have trouble processing it.[23–25] Moreover, information that is inconsistent with one's beliefs produces a negative affective response, as shown in research on cognitive consistency since the 1950s.[26,27] Either of these experiences can signal that something does not feel right, which may prompt more critical thought and analysis.

   In contrast, when the new information matches one's beliefs, processing is easy, and people tend to nod along. As an example, suppose you are asked, "How many animals of each kind did Moses take on the ark?" Most people answer "two" despite knowing that the biblical actor was Noah, not Moses[28]—the biblically themed question feels familiar, and people focus on

what they are asked about (how many?) rather than the background details (who). But when the question is printed in a difficult-to-read font that impedes easy processing, the words do not flow as smoothly. Now something seems to feel wrong, and more people notice the error embedded in the question.[29]

4. Coherence: Does It Tell a Good Story?

When details are presented as part of a narrative and individual elements fit together in a coherent frame, people are more likely to think it is true.[30,31] For instance, in a 1992 article about juror decision making, Nancy Pennington and Reid Hastie of the University of Colorado described experiments in which they asked volunteers to render verdicts after reading transcripts of cases consisting of several witness statements. The researchers varied the way information was presented: Either evidence was blocked so that all of the evidence (across several witnesses) regarding motive appeared as a summary, or it was presented more like a story, as witness narratives. The researchers found that people tended to believe the witnesses more when the same evidence was presented in the format of a coherent story.

   In fact, when asked to remember a story, people often remember it in ways that make it more coherent, even filling in gaps and changing elements.[32] Maryanne Garry of the University of Wellington in New Zealand and her colleagues had volunteers watch a video of a woman making a sandwich. Although participants probably thought they saw the whole video, certain parts of the sandwich-making process were not shown. In a later memory test, participants confidently but falsely remembered events they had never witnessed in the video.

   When a story feels coherent, people think that it makes more sense, and they enjoy reading it more.[33,34] Coherent stories flow more smoothly and are easier to process than incoherent stories with internal contradictions are.[30] There are several ways to increase the chances that readers will feel as though they are reading a coherent story. For example, in one line of studies, Jonathan Leavitt and Nicholas Christenfeld of the University of California, San Diego, gave some participants summary information that enabled them to anticipate a story's ending before they began to read it. After reading, those who had the extra information said they enjoyed the story more—having some prior context lent the story more coherence and made it easier to follow.

5. Credibility: Does It Come from a Credible Source?

Not surprisingly, people are more likely to accept information from a credible source than from a less credible one.[35,36] People evaluate the credibility of a source in many ways, such as by looking at the source's expertise, past statements, and likely motives. Alternatively, people can again consult their feelings about the source. When they do so, the apparent familiarity of the source looms large. Repeatedly seeing a face is enough to increase perceptions of honesty, sincerity, and general agreement with what that person says.[37,38] Even the ease of pronouncing the speaker's name influences credibility: When a person's name is easy to say, people are more likely to believe what they hear from the person.[39] Thus, a source can seem credible simply because the person feels familiar.

An exception to this rule is when people realize that the person seems familiar for a bad reason. For example, although the name Adolf Hitler is familiar and easy to pronounce, it does not lend credibility. Similarly, familiarity is unlikely to enhance the credibility of a source that is closely identified with a view that one strongly opposes, as might happen if the source is a politician from an opposing party. (See the sidebar *Political Messages from the Other Side*.) In these cases, familiarity with the source comes with additional information that serves as a warning signal and prompts closer scrutiny.

A source also seems more credible when the message is easy to process. For example, people are more likely to believe statements when they are made in a familiar and easy-to-understand accent rather than a difficult-to-understand one. In a 2010 study, for instance, Shin Lev-Ari and Boaz Keysar of the University of Chicago asked native speakers of American English to rate the veracity of trivia statements (such as "A giraffe can go longer without water than a camel can"). Volunteers rated statements recited by native English speakers more truthful than statements recited by speakers of accented English (whose native tongues included Polish, Turkish, Italian, and Korean).[40]

Summary of Truth Evaluation

Regardless of which truth criteria people draw on, easily processed information enjoys an advantage over information that is difficult to process: It feels more familiar, widely held, internally consistent, compatible with one's

---

Political Messages from the Other Side

Messages from the other side of a political debate rarely change partisan minds. The five truth tests discussed in the main text shed some light on why. To begin with, a message from a political opponent comes from a source that one has already identified as being associated with other interests, thus limiting its credibility. Moreover, its content is likely to be at odds with several of one's beliefs. Accordingly, thinking of many arguments that support a message from the other side is difficult, but coming up with many counterarguments is easy. In addition, opposing beliefs interfere with the processing of the information, so arguments will not seem to flow smoothly. This limits the perceived coherence of the message—it is just not a good story. Finally, one's own social network is unlikely to agree with other-side messages, thus limiting perceived social consensus as well.

As a result, messages that contradict a person's worldview and advocate opposing positions are unlikely to feel true and compelling to that person. This effect is not just evidence for the stubbornness of partisans but inherent in how people gauge truth: The dominant truth criteria inherently place beliefs of the other side at a disadvantage.

However, the other side's messages may gain in acceptance as time passes. For example, election campaigns expose all citizens to messages that are closely linked to partisan sources. Yet, as time goes by, the specific source will be forgotten, but the message may feel fluent and familiar when it is encountered after the campaign is over. That is, although one may reject a message from the other side at first, the message itself may seem more plausible later on, when the original source cannot be remembered. At that point, it may receive less scrutiny, and people may nod along because of the fluency resulting from previous encounters.

---

beliefs, and likely to have come from a credible source. In short, easy processing gives folks an intuitive feeling of believability and helps pass the Big Five major truth criteria tests outlined above.[7] Put simply, when thought flows smoothly, people tend to accept them without analyzing them too closely.

Alternatively, information that is difficult to process, feels unfamiliar, and makes people stumble is more likely to trigger critical analysis. When something feels wrong, people pay closer attention, look for more relevant information, and are willing to invest more effort into figuring out what is likely to be true. People are

Fluency: When It Is Easy, It Seems Familiar, and Familiar Feels True

Any mental act, from reading and hearing to remembering and evaluating, can feel easy or difficult. Material that is easy to process feels fluent, in contrast to material that is difficult to process, which may make the reader stumble. People are sensitive to these feelings but not to where they come from. For example, familiar material is easier to read than unfamiliar material is, but not everything that is easy to read is also familiar.

Many things can influence the feeling of fluency. Influences include presentation characteristics, such as print font, color contrast, or a speaker's accent, and content characteristics, such as the complexity and flow of an argument. They also include the receiver's expertise and history with the material, such as how often one has seen it before and how long ago one saw it.

When any of these factors make processing easy, they increase the likelihood that a message is accepted as true. Hence, people are more likely to consider a statement true when it is presented, for example, in high color contrast, in a more simple font or in a rhyming form.[A,B]

More likely to be judged true:	*Less likely to be judged true:*
Orsono is a city in Chile	Orsono is a city in Chile
Orsono is a city in Chile	Orsono is a city in Chile
Woes unite foes	Woes unite enemies

A. Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8,* 338–342.
B. McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science, 11,* 424–428.

also more likely to notice misleading questions and to critically examine their own beliefs.[7,29,41] If their critical analysis reveals something faulty, they will reject the message. But if the arguments hold up to scrutiny, a message that initially felt wrong may end up being persuasive. Nevertheless, in most cases, recipients will conclude that a message that feels wrong is not compelling. After all, at first glance, it did not meet the Big Five truth criteria discussed above.

Repeating False Information: A Bad Idea

The reviewed research sheds light on why some correction strategies may unintentionally cement the ideas they are trying to correct: When a correction attempt increases the ease with which the false claim can be processed, it also increases the odds that the false claim feels true when it is encountered again at a later point in time.

Repetition Increases Acceptance

The popular strategy of juxtaposing myths and facts necessarily involves a repetition of the false claims

(or *myths*) in order to confront them with the facts. A growing number of studies show that this strategy can have unintended consequences: increasing the acceptance of false beliefs, spreading them to new segments of the population, and creating the perception that the false beliefs are widely shared. For example, in a 2005 study,[42] Ian Skurnik of the University of Toronto and his colleagues had participants view health-related statements. They told them which ones were true and which were false. When participants were tested immediately, they were able to recall this information from memory and could distinguish fact from fiction. But 3 days later, after their memories had a chance to fade, participants were more likely to think that any statement they had seen was true, whether it had been presented as true or false. Moreover, the acceptance of false statements increased with the number of warnings: Participants who had been told thrice that a statement was false were more likely to accept it as true than were those who had only been told once. Older participants were particularly vulnerable to this bias, presumably because their poorer memory made it harder to remember the details of what they had heard earlier.

> *Participants who had been told thrice that a statement was false were more likely to accept it as true than were those who had only been told once.*

Startlingly, it takes neither 3 days nor old age for such a paradoxical effect to occur. When undergraduates viewed a myths-and-facts flyer about the flu taken from the CDC website, they remembered some myths as facts after only 30 minutes.[6] Moreover, despite the flyer's promotion of the flu vaccine for their age group, participants who had read the myths-and-facts flyer reported lower intentions to get a flu vaccination than did participants who read only the facts. Worse, their reported intentions to get vaccinated were even lower than those of control participants who had not been exposed to any message about the flu. Apparently, realizing there might be some controversy about the issue was sufficient to undermine healthy intentions.

Repetition Spreads Misinformation to New Audiences

Myths typically take root in a small segment of the population, yet sometimes a myth breaks free and spreads to larger audiences. Ironically, the cause of the spread may be education campaigns. Although one may hope that the clear juxtaposition of myth and fact teaches the new audience what's right and wrong and inoculates them against later misinformation, this is not always the case. Instead, a well-intentioned information campaign may have the unfortunate effect of spreading false beliefs to a broader population.

The flesh-eating bananas rumor is an example. It moved from the fringes of the Internet to mainstream media after the CDC published its correction, which was picked up by the *Los Angeles Times*. After a while, people misremembered the sources of the correction as the sources of the false information itself, resulting in the impression that flesh-eating bananas are a real problem.[43] This retrospective attribution of a myth to a more credible source goes beyond the more common observation that messages initially seen as unconvincing because they come from an untrustworthy source can exert an influence later on, once their source is forgotten (a phenomenon known as the *sleeper effect*).[44,45]

Myth-Busting Can Convey Controversy

The popular myth-versus-fact formats also convey the impression that a significant number of people hold a different position or positions on an issue, or else there would be no reason to juxtapose myths and facts. So although the myth-versus-fact format may increase readership and engagement, it also can make a topic seem controversial and render the truth unclear. It tells people that either side could be right and can make a vocal minority seem larger than it is. People with limited expertise in an area are therefore likely to defer judgment and hesitate to take sides. This is particularly likely in scientific controversies, where the facts are difficult for the public to evaluate, as is the case with certain dietary approaches or health treatments[4] as well as for climate change.[13,46] The strategy of emphasizing controversy to engage readers is problematic when the actual facts have been well demonstrated, because it undermines the credibility of the facts and facilitates overestimates of the disagreement.

Anecdotes and Photographs Reinforce the Message

Anecdotes and photos serve several communicative goals—they capture attention, boost comprehension, and enhance the readability of associated text.[47–49] This makes the content easier to imagine, which can artificially boost its perceived truth.[50]

Anecdotes promote understanding because they link new information with prior knowledge and evoke vivid pictures in people's minds. For these reasons, they can have powerful effects on people's beliefs, leading them to ignore available statistics and scientific facts and use feelings and intuition as measures by which to evaluate information. In 2005, Angela Fagerlin, now at the University of Michigan, and her colleagues asked study volunteers to read a scenario about angina and to choose between bypass surgery and balloon angioplasty. They tended to overlook statistical data about the cure rates and instead choose the option that included anecdotes of those who underwent that procedure.[51]

Photos can produce similar effects, even when they have no probative value for the claim with which they are paired. In one experiment conducted by Eryn Newman of the University of Southern California and colleagues,[50] participants in New Zealand were shown

Photographs and Truthiness

Messages or claims that appear with photos catch the eye and generally are more easily understood and remembered. But adding a photo to claims can also add authority: People are more likely to think claims are true when they appear with a photo. Photos have this influence even when they provide no probative evidence about whether the claim is correct. For instance, people are more likely to believe the claim "Magnesium is the liquid metal inside a thermometer" when they see a photo of a thermometer, even one that provides no information regarding what metal can be found inside. (Most household glass thermometers use alcohol with red dye.) One reason why photos bring about this truthiness effect is that they make it easy for the reader to understand and imagine the claim. As a result, the claim feels fluent, familiar, and true.

Want to convince people that Nick Cave is dead or Nick Cave is alive? Easy. Just add his picture to either claim and voila! People believe.

(For more information on the experiment that investigated this scenario, see "Nonprobative Photographs (or Words) Inflate Truthiness," by E. J. Newman, M. Garry, D. M. Bernstein, J. Kantner, and D. S. Lindsay, 2012, *Psychonomic Bulletin & Review, 19,* 969–974.)

a picture of Nick Cave, a musician with the Australian band the Bad Seeds. When the photo accompanied the claim "Nick Cave is alive," people were more likely to agree that he is, indeed, alive than when no photo was presented. But the same photo also made people more likely to think that Nick Cave is dead when the photo accompanied the claim "Nick Cave is dead." (For the record, Nick Cave is alive as of this writing.)

Other more superficial communication approaches can produce similar effects. For example, rhyme can enhance memory for material by serving as a mnemonic device. But rhyme can also enhance the credibility of a message, even if it does not add any supporting evidence, by making words flow smoothly. In 2000, Mathew McGlone and Jessica Tofighbakhsh, then

of Lafayette College, asked study participants to evaluate sayings about human behavior and rate the truth of each saying. When the sayings rhymed (for example, "Woes unite foes"), people were more likely to think they were true representations of human conduct than when the sayings did not ("Woes unite enemies").[52]

In sum, anecdotes, pictures, and rhymes that contain little informational value are usually offered to engage readers. But they can nevertheless influence outcomes because they scaffold mental imagery, increase the ease with which a message is processed, produce a feeling of remembering, and systematically bias people to believe information whether it is true or false. For that reason, these communication devices can thwart the intended educational effect when they are presented with false information; we therefore discourage their use when written content contains myths or retractions.

Key Communication Strategies for Making Truths Stick and Myths Fade

So how can one correct false beliefs and increase public knowledge without propagating misinformation? The available research indicates that information is more likely to stick the more easily it can be processed and the more familiar it feels. Accordingly, the overarching goal for any communication strategy is to increase the fluency and familiarity of correct information and to decrease the fluency and familiarity of misinformation. Attempts at correcting misinformation—for example, using the myth-versus-fact setup—often fail because they center on the false information and unintentionally increase the ease with which false information can be processed when it is encountered again. Increasing the fluency and familiarity of true information can be achieved in three key ways.

The first way is through repetition—specifically, repetition of the correct information, not the misinformation one wants to undermine. For this reason, it is usually better to ignore false information than to repeat it. The popular myth-versus-fact format unwittingly reinforces the myths by repeating them, which makes them more influential once memory for the less familiar (and often more complex) facts fades. Focus rather on the facts, making them easy to understand and easy to remember. Instead of repeating various vaccination myths, for example, a more effective strategy is to document why vaccinations are safe and to emphasize

the scientific evidence that vaccines promote health and not harm.

> *Repeat correct information, not the misinformation one wants to undermine.*

Sometimes there are legal requirements to repeat false information in the context of a correction. In such cases, it is important to provide a fluent and coherent account of why the false information was presented to begin with. Consider the myth that autism is caused by childhood vaccines. A straightforward, easy-to-comprehend account of how the discovery of an alleged autism–vaccine link was completely made up and based on fraudulent data that cost the principal author his professional license will be more effective in addressing the misinformation than simply labeling the original myth *discredited,* as many news outlets routinely do.

Second, true information needs to be made as accessible as possible. Unfortunately, the truth is often more complicated than the myth, which usually involves considerable simplification. This puts the truth at a disadvantage because it is harder to process, understand, and remember. Presenting true information in ways that make processing it as easy as possible is therefore important. This requires clear, step-by-step exposition and the avoidance of jargon. Other more cosmetic changes can also make the truth easily digestible—choosing an easy-to-read font and ensuring the speaker's pronunciation is easy to understand can increase the fluency of a message. It also helps when the true information is accompanied by pictures that make the information easy to imagine or when key parts of the repeated message rhyme.

Finally, at the individual level, one of the most powerful strategies for avoiding misinformation is to know it is coming.[5] In one study, Stephan Lewandowsky of the University of Bristol and his colleagues asked participants to read a short description about a bus accident. After reading the passage, participants were told that some of the information was wrong. Despite the retractions, many participants held on to the inaccurate details that they learned from the initial description of the bus accident. That is, once the story was told, it was difficult to cleave out inaccuracies.

Two strategies can effectively prevent such misconceptions. One is to provide accurate details that present an alternative account of the misinformation, increasing the chances of people remembering the true information and allowing the false details to fade away. The second is to warn people before they read the passage about the influence of misinformation. Pre-exposure warnings can alert people to carefully scrutinize the content of information and ward off false details.[53–56]

Although research shows that warnings are more efficient when they are received prior to the false information, this is not where they are commonly placed. In the health domain, the law requires that labels claiming unsubstantiated health benefits must include a disclaimer: "This product is not intended to diagnose, treat, cure, or prevent any disease."[57] Such disclaimers commonly follow the unsubstantiated claims. Moving them to the top of a label or the beginning of radio advertisement is likely to enhance their impact.

In sum, the available research shows that highlighting false information and then attempting to unwind its effects is usually a bad idea. More promising communication strategies focus on the truth, making it easier to process and more handily remembered, which increases the chance that the correct message sticks.

author affiliation

Schwarz, Department of Psychology, University of Southern California; Newman, Dornsife Mind and Society Center, University of Southern California; Leach, Sol Price School of Public Policy, University of Southern California. Corresponding author's e-mail: norbert.schwarz@usc.edu

References

1. Fragale, A. R., & Heath, C. (2004). Evolving informational credentials: The (mis)attribution of believable facts to credible sources. *Personality and Social Psychology Bulletin, 30,* 225–236.
2. Kata, A. (2010). A postmodern Pandora's box: Antivaccination misinformation on the Internet. *Vaccine, 28,* 1709–1716.
3. Goldacre, B. (2009). Media misinformation and health behaviours. *Lancet Oncology, 10,* 848.
4. Ayoob, K. T., Duyff, R. L., & Quagliani, D. (2002). Position of the American Dietetic Association: Food and nutrition misinformation. *Journal of the American Dietetic Association, 102,* 260–266.
5. Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*(3), 106–131.

6. Schwarz, N., Sanna, L. J., Skurnik, I., & Yoon, C. (2007). Metacognitive experiences and the intricacies of setting people straight: Implications for debiasing and public information campaigns. *Advances in Experimental Social Psychology, 39,* 127–161.

7. Schwarz, N. (2015). Metacognition. In M. Mikulincer, P. R. Shaver, E. Borgida, & J. A. Bargh (Eds.), *APA handbook of personality and social psychology: Attitudes and social cognition* (Vol. 1, pp. 203–229). Washington, DC: American Psychological Association.

8. Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7,* 117–140.

9. Newcomb, T. M. (1943). *Personality and social change: Attitude formation in a student community.* Fort Worth, TX: Dryden Press.

10. Visser, P. S., & Mirabile, R. R. (2004). Attitudes in the social context: The impact of social network composition on individual-level attitude strength. *Journal of Personality and Social Psychology, 87,* 779–795.

11. Harris, A. J. L., & Hahn, U. (2009). Bayesian rationality in evaluating multiple testimonies: Incorporating the role of coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1366–1372.

12. Ross, M., Buehler, R., & Karr, J. W. (1998). Assessing the accuracy of conflicting autobiographical memories. *Memory & Cognition, 26,* 1233–1244.

13. Lewandowsky, S., Gignac, G. E., & Vaughan, S. (2013). The pivotal role of perceived scientific consensus in acceptance of science. *Nature Climate Change, 3,* 399–404.

14. Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology, 92,* 821–833.

15. Foster, J. L., Huthwaite, T., Yesberg, J. A., Garry, M., & Loftus, E. F. (2012). Repetition, not number of sources, increases both susceptibility to misinformation and confidence in the accuracy of eyewitnesses. *Acta Psychologica, 139,* 320–326.

16. Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of Personality and Social Psychology, 64,* 211–220.

17. Schwarz, N., Bless, H., Strack, F., Klumpp, G., Rittenauer-Schatka, H., & Simons, A. (1991). Ease of retrieval as information: Another look at the availability heuristic. *Journal of Personality and Social Psychology, 61*, 195–202.

18. Larrick, R. P. (2004). Debiasing. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 316–337). Oxford, United Kingdom: Blackwell.

19. Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 497–502.

20. Abelson, R. P. (1968). Theories of cognitive consistency: A sourcebook. Chicago, IL: Rand McNally.

21. McGuire, W. J. (1972). Attitude change: An information processing paradigm. In C. G. McClintock (Ed.), *Experimental social psychology* (pp. 108–141). New York, NY: Holt, Rinehart and Winston.

22. Wyer, R. S. (1974). *Cognitive organization and change: An information processing approach.* Potomac, MD: Erlbaum.

23. Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology, 71,* 5–24.

24. Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50,* 755–769.

25. Winkielman, P., Huber, D. E., Kavanagh, L., & Schwarz, N. (2012). Fluency of consistency: When thoughts fit nicely and flow smoothly. In B. Gawronski & F. Strack (Eds.), *Cognitive consistency: A fundamental principle in social cognition* (pp. 89–111). New York, NY: Guilford Press.

26. Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

27. Gawronski, B., & Strack, F. (2012). *Cognitive consistency: A fundamental principle in social cognition.* New York, NY: Guilford Press.

28. Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior, 20,* 540–551.

29. Song, H., & Schwarz, N. (2008). Fluency and the detection of distortions: Low processing fluency attenuates the Moses illusion. *Social Cognition, 26,* 791–799.

30. Johnson-Laird, P. N. (2012). Inference with mental models. In K. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 134–145). New York, NY: Oxford University Press.

31. Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62,* 189–206.

32. Gerrie, M. P., Belcher, L. E., & Garry, M. (2006). 'Mind the gap': False memories for missing aspects of an event. *Applied Cognitive Psychology, 20,* 689–696.

33. Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

34. Leavitt, J., & Christenfeld, N. J. (2013). The fluency of spoilers: Why giving away endings improves stories. *Scientific Study of Literature, 3,* 93–104.

35. Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes.* Orlando, FL: Harcourt Brace Jovanovich College.

36. Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change.* New York, NY: Springer.

37. Brown, A. S., Brown, L. A., & Zoccoli, S. L. (2001). Repetition-based credibility enhancement of unfamiliar faces. *The American Journal of Psychology, 115,* 199–209.

38. Weisbuch, M., & Mackie, D. (2009). False fame, perceptual clarity, or persuasion? Flexible fluency attribution in spokesperson familiarity effects. *Journal of Consumer Psychology, 19,* 62–72.

39. Newman, E. J., Sanson, M., Miller, E. K., Quigley-McBride, A., Foster, J. L., Bernstein, D. M., & Garry, M. (2014). People with easier to pronounce names promote truthiness of claims. *PloS One, 9*(2), Article e88671. doi:10.1371/journal.pone.0088671

40. Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46,* 1093–1096.

41. Lee, D. S., Kim, E., & Schwarz, N. (2015). Something smells fishy: Olfactory suspicion cues improve performance on the Moses illusion and Wason rule discovery task. *Journal of Experimental Social Psychology, 59,* 47–50.

42. Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research, 31,* 713–724.

43. Emery, D. (2000,). *The great Internet banana scare of 2000:* "Killer flesh-eating bananas" rumor floods Internet. Retrieved August 2, 2016, from: http://urbanlegends.about.com/od/fooddrink/a/killer_bananas.htm

44. Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15,* 635–650. doi:10.1086/266350

45. Pratkanis, A. R., Greenwald, A. G., Leippe, M. R., & Baumgardner, M. H. (1988). In search of reliable persuasion effects: III. The sleeper effect is dead: Long live the sleeper

effect. *Journal of Personality and Social Psychology, 54,* 203–218. doi:10.1037/0022-3514.54.2.203

46. Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Seepage: Climate change denial and its effect on the scientific community. *Global Environmental Change, 33,* 1–13.

47. Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling, 61,* 173–190.

48. Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88,* 49–63.

49. Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63,* 760–769.

50. Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review, 19,* 969–974.

51. Fagerlin, A., Wang, C., & Ubel, P. A. (2005). Reducing the influence of anecdotal reasoning on people's health care decisions: Is a picture worth a thousand statistics? *Medical Decision Making, 25,* 398–405.

52. McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science, 11,* 424–428.

53. Blank, H., & Launay, C. (2014). How to protect eyewitness memory against the misinformation effect: A meta-analysis of post-warning studies. *Journal of Applied Research in Memory and Cognition, 3,* 77–88.

54. Butler, A. C., Zaromb, F. M., Lyle, K. B., & Roediger, H. L., III. (2009). Using popular films to enhance classroom learning: The good, the bad, and the interesting. *Psychological Science, 20,* 1161–1168.

55. Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition, 38,* 1087–1100.

56. Tousignant, J. P., Hall, D., & Loftus, E. F. (1986). Discrepancy detection and vulnerability to misleading postevent information. *Memory & Cognition, 14,* 329–338.

57. Certain Types of Statements for Dietary Supplements, 21 C.F.R. § 101.93 (2015).

Behavioral Science & Policy (BSP) is an international, peer-reviewed publication of the Behavioral Science & Policy Association and Brookings Institution Press. BSP features short, accessible articles describing actionable policy applications of behavioral scientific research that serves the public interest. Articles submitted to BSP undergo a dual-review process: For each article, leading disciplinary scholars review for scientific rigor and experts in relevant policy areas review for practicality and feasibility of implementation. Manuscripts that pass this dual-review are edited to ensure their accessibility to policy makers, scientists, and lay readers. BSP is not limited to a particular point of view or political ideology.

Manuscripts can be submitted in a number of different formats, each of which must clearly explain specific implications for public- and/or private-sector policy and practice.

External review of the manuscript entails evaluation by at least two outside referees—at least one in the policy arena and at least one in the disciplinary field.

Professional editors trained in BSP's style work with authors to enhance the accessibility and appeal of the material for a general audience.

Each of the sections below provides general information for authors about the manuscript submission process. We recommend that you take the time to read each section and review carefully the BSP Editorial Policy before submitting your manuscript to *Behavioral Science & Policy*.

Manuscript Categories
Manuscripts can be submitted in a number of different categories, each of which must clearly demonstrate the empirical basis for the article as well as explain specific implications for (public and/or private-sector) policy and practice:

- Proposals (≤ 2,500 words) specify scientifically grounded policy proposals and provide supporting evidence including concise reports of relevant studies. This category is most appropriate for describing new policy implications of previously published work or a novel policy recommendation that is supported by previously published studies.
- Reports (≤ 3000 words) provide a summary of output and actionable prescriptions that emerge from a workshop, working group, or standing organization in the behavioral policy space. In some cases such papers may consist of summaries of a much larger published report that also includes some novel material such as meta-analysis, actionable implications, process lessons, reference to related work by others, and/or new results not presented in the initial report. These papers are not merely summaries of a published report, but also should provide substantive illustrations of the research or recommendations and insights about the implications of the report content or process for others proposing to do similar work. Submitted papers will undergo BSP review for rigor and accessibility that is expedited to facilitate timely promulgation.

- Findings (≤ 4,000 words) report on results of new studies and/or substantially new analysis of previously reported data sets (including formal meta-analysis) and the policy implications of the research findings. This category is most appropriate for presenting new evidence that supports a particular policy recommendation. The additional length of this format is designed to accommodate a summary of methods, results, and/or analysis of studies (though some finer details may be relegated to supplementary online materials).
- Reviews (≤ 5,000 words) survey and synthesize the key findings and policy implications of research in a specific disciplinary area or on a specific policy topic. This could take the form of describing a general-purpose behavioral tool for policy makers or a set of behaviorally grounded insights for addressing a particular policy challenge.
- Other Published Materials. BSP will sometimes solicit or accept *Essays* (≤ 5,000 words) that present a unique perspective on behavioral policy; *Letters* (≤ 500 words) that provide a forum for responses from readers and contributors, including policy makers and public figures; and *Invitations* (≤ 1,000 words with links to online Supplemental Material), which are requests from policy makers for contributions from the behavioral science community on a particular policy issue. For example, if a particular agency is facing a specific challenge and seeks input from the behavioral science community, we would welcome posting of such solicitations.

Review and Selection of Manuscripts
On submission, the manuscript author is asked to indicate the most relevant disciplinary area and policy area addressed by his/her manuscript. (In the case of some papers, a "general" policy category designation may be appropriate.) The relevant Senior Disciplinary Editor and the Senior Policy Editor provide an initial screening of the manuscripts. After initial screening, an appropriate Associate Policy Editor and Associate Disciplinary Editor serve as the stewards of each manuscript as it moves through the editorial process. The manuscript author will receive an email within approximately two weeks of submission, indicating whether the article has been sent to outside referees for further consideration. External review of the manuscript entails evaluation by at least two outside referees. In most cases, Authors will receive a response from BSP within approximately 60 days of submission. With rare exception, we will submit manuscripts to no more than two rounds of full external review. We generally do not accept re-submissions of material without an explicit invitation from an editor. Professional editors trained in the BSP style will collaborate with the author of any manuscript recommended for publication to enhance the accessibility and appeal of the material to a general audience (i.e., a broad range of behavioral scientists, public- and private-sector policy makers, and educated lay public). We anticipate no more than two rounds of feedback from the professional editors.

Standards for Novelty

BSP seeks to bring new policy recommendations and/or new evidence to the attention of public and private sector policy makers that are supported by rigorous behavioral and/or social science research. Our emphasis is on novelty of the policy application and the strength of the supporting evidence for that recommendation. We encourage submission of work based on new studies, especially field studies (for Findings and Proposals) and novel syntheses of previously published work that have a strong empirical foundation (for Reviews).

BSP will also publish novel treatments of previously published studies that focus on their significant policy implications. For instance, such a paper might involve re-working of the general emphasis, motivation, discussion of implications, and/or a re-analysis of existing data to highlight policy-relevant implications or prior work that have not been detailed elsewhere.

In our checklist for authors we ask for a brief statement that explicitly details how the present work differs from previously published work (or work under review elsewhere). When in doubt, we ask that authors include with their submission copies of related papers. Note that any text, data, or figures excerpted or paraphrased from other previously published material must clearly indicate the original source with quotation and citations as appropriate.

Authorship

Authorship implies substantial participation in research and/or composition of a manuscript. All authors must agree to the order of author listing and must have read and approved submission of the final manuscript. All authors are responsible for the accuracy and integrity of the work, and the senior author is required to have examined raw data from any studies on which the paper relies that the authors have collected.

Data Publication

BSP requires authors of accepted empirical papers to submit all relevant raw data (and, where relevant, algorithms or code for analyzing those data) and stimulus materials for publication on the journal web site so that other investigators or policymakers can verify and draw on the analysis contained in the work. In some cases, these data may be redacted slightly to protect subject anonymity and/or comply with legal restrictions. In cases where a proprietary data set is owned by a third party, a waiver to this requirement may be granted. Likewise, a waiver may be granted if a dataset is particularly complex, so that it would be impractical to post it in a sufficiently annotated form (e.g. as is sometimes the case for brain imaging data). Other waivers will be considered where appropriate. Inquiries can be directed to the BSP office.

Statement of Data Collection Procedures

BSP strongly encourages submission of empirical work that is based on multiple studies and/or a meta-analysis of several datasets. In order to protect against false positive results, we ask that authors of empirical work fully disclose relevant details concerning their data collection practices (if not in the main text then in the supplemental online materials). In particular, we ask that authors report how they determined their sample size, all data exclusions (if any), all manipulations, and all measures

in the studies presented. (A template for these disclosures is included in our checklist for authors, though in some cases may be most appropriate for presentation online as Supplemental Material; for more information, see Simmons, Nelson, & Simonsohn, 2011, *Psychological Science, 22, 1359–1366*).

Copyright and License

Copyright to all published articles is held jointly by the Behavioral Science & Policy Association and Brookings Institution Press, subject to use outlined in the *Behavioral Science & Policy* publication agreement (a waiver is considered only in cases where one's employer formally and explicitly prohibits work from being copyrighted; inquiries should be directed to the BSPA office). Following publication, the manuscript author may post the accepted version of the article on his/her personal web site, and may circulate the work to colleagues and students for educational and research purposes. We also allow posting in cases where funding agencies explicitly request access to published manuscripts (e.g., NIH requires posting on PubMed Central).

Open Access

BSP posts each accepted article on our website in an open access format at least until that article has been bundled into an issue. At that point, access is granted to journal subscribers and members of the Behavioral Science & Policy Association. Questions regarding institutional constraints on open access should be directed to the editorial office.

Supplemental Material

While the basic elements of study design and analysis should be described in the main text, authors are invited to submit Supplemental Material for online publication that helps elaborate on details of research methodology and analysis of their data, as well as links to related material available online elsewhere. Supplemental material should be included to the extent that it helps readers evaluate the credibility of the contribution, elaborate on the findings presented in the paper, or provide useful guidance to policy makers wishing to act on the policy recommendations advanced in the paper. This material should be presented in as concise a manner as possible.

Embargo

Authors are free to present their work at invited colloquia and scientific meetings, but should not seek media attention for their work in advance of publication, unless the reporters in question agree to comply with BSP's press embargo. Once accepted, the paper will be considered a privileged document and only be released to the press and public when published online. BSP will strive to release work as quickly as possible, and we do not anticipate that this will create undue delays.

Conflict of Interest

Authors must disclose any financial, professional, and personal relationships that might be construed as possible sources of bias.

Use of Human Subjects

All research using human subjects must have Institutional Review Board (IRB) approval, where appropriate.

sponsors

The Behavioral Science & Policy Association is grateful to the sponsors and partners who generously provide continuing support for our non-profit organization.

Alfred P. Sloan FOUNDATION

BROOKINGS

Duke SOCIAL SCIENCE RESEARCH INSTITUTE

MacArthur Foundation

UCLAAnderson SCHOOL of MANAGEMENT

THE WILLIAM AND FLORA HEWLETT FOUNDATION

To become a Behavioral Science & Policy Association sponsor, please contact BSPA at bspa@behavioralpolicy.org or 1-919-681-5932.

behavioral
science & policy

where behavioral research meets policy + practice

who we are

the Behavioral Science & Policy Association is a global hub of behavioral science resources, curated by leading scholars and policymakers, aimed at facilitating positive change and innovative solutions to a range of societal changes.

our mission

to foster and connect a growing community of interdisciplinary practitioners, providing thoughtful application of rigorous behavioral science research for the public and private sectors, with a simple goal in mind: addressing social change for the public interest.

membership

there is a growing movement among social scientists and leaders within the public and private sector, dedicated to grounding important decisions in strong scientific evidence

BSPA plays a key role in this movement, encouraging decisions to be based on evidence. we need you to join us in the effort to make a lasting impact.

as a BSPA member, you will receive numerous benefits including an online subscription to *Behavioral Science & Policy*, early-bird rates for conferences, workshops and briefings, exclusive access to BSPA online webinars and podcasts, waived fees for journal submissions and more.

be a leader in our drive for change at https://behavioralpolicy.org/signup

call for submissions

Behavioral Science & Policy is an international, peer-reviewed journal featuring succinct and accessible articles outlining actionable policy applications of behavioral scientific research that serve the public interest.

BSP journal submissions undergo a dual-review process. leading scholars from specific disciplinary areas review articles to assess their scientific rigor; while at the same time, experts in relevant policy areas evaluate these submissions for relevance and feasibility of implementation.

manuscripts that pass this dual-review are edited to ensure accessibility to scientists, policymakers, and lay readers. BSPA is not limited to a particular point of view or political ideology, and is a publication of the Behavioral Science & Policy Association and the Brookings Institution Press

we encourage you to submit your manuscript today to *Behavioral Science & Policy*, at http://behavioralpolicy.org/journal