



The Power of a Penny

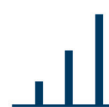
Building Knowledge to Invest in What Works in Education

Robert Balfanz

Center for Social Organization of Schools

Johns Hopkins School of Education

A Part of the Invest in
What Works Policy Series

 **RESULTS**
★★★
FOR AMERICA

An Initiative of *America Achieves*

The Power of a Penny

Building Knowledge to Invest in What Works in Education

Robert Balfanz

Center for Social Organization of Schools
Johns Hopkins School of Education

EVERYONE CENTER
GRADUATES

 **RESULTS**
★★★ **FOR AMERICA**
An Initiative *of* America Achieves

Invest in What Works Policy Series

This report is part of Results for America’s “Invest in What Works Policy Series,” which provides ideas and supporting research to policymakers to drive public funds toward evidence-based, results-driven solutions. The series includes policy reports, expert roundtable discussions, and public events. Bloomberg Philanthropies graciously provided support for the launch of the Invest in What Works Policy Series.

About the Everyone Graduates Center

The Everyone Graduates Center at the School of Education at Johns Hopkins University offers research on, and a systematic and comprehensive approach to identifying, the nation’s dropout crisis and its remedies; effective tools and models designed to keep students engaged, achieving and advancing; capacity-building efforts to enable states, communities, school districts, and schools to provide all their students with the support they need to succeed. The center also provides consulting services and technical assistance on early warning indicator systems and other strategies designed to raise the nation’s graduation rate. www.every1graduates.org

About Results for America

Results for America, an initiative of America Achieves, is committed to improving outcomes for young people, their families and communities by shifting public resources toward programs and practices that use evidence and data to improve quality and get better results.

**Bloomberg
Philanthropies**

*This report has been published with the
generous support of Bloomberg Philanthropies.*

CONTENTS

EXECUTIVE SUMMARY	3
INTRODUCTION.....	5
What is Moneyball for Government?.....	5
WHY WE NEED TO INVEST MORE IN EVALUATIONS OF FEDERALLY FUNDED EDUCATION PROGRAMS	8
THE U.S. DEPARTMENT OF EDUCATION NEEDS TO LEAD THE WAY	9
Current Approach to Program Evaluation.....	9
Movement in the Right Direction.....	10
LEVERAGE EVALUATION TO MAXIMIZE IMPACT: A STRATEGIC FUND TO INFORM POLICY AND PRACTICE.....	12
POLICY RECOMMENDATIONS.....	13
CONCLUSION.....	23
REFERENCES.....	24
Acknowledgement.....	25

EXECUTIVE SUMMARY

The federal government spends less than one-tenth of a penny, out of every dollar invested in education programs, to evaluate those programs. That ratio is far lower than other fields, and it means that policymakers often lack basic evidence on how to spend taxpayer resources wisely to improve education.

This paper outlines why and how the U.S. Department of Education should leverage a 1% investment in evaluation—a penny on the dollar—to develop and implement a comprehensive evaluation system. Such a system could significantly increase the impact and cost-effectiveness of federal investment in education, and build the knowledge base to help teachers and school leaders solve chronic education problems.

In May 2014, Results for America (RFA) issued its third [federal Investing in What Works Index](#), which highlighted the extent to which the U.S. Department of Education and other federal agencies are building the infrastructure needed to use evidence and evaluation in budget, policy, and management decisions (RFA 2014). The Department scored well on the index, but it has plenty of room for improvement, starting with investing 1% of funds in rigorous, independent evaluations.

This paper makes eight recommendations for how the Department could use evaluation funding to build the knowledge base and improve its programs:

1. Use scientifically rigorous methods to evaluate programs and build a cumulative evidence base.
2. Focus program evaluations on finding what works—where, when, how, and under what circumstances—rather than a thumbs-up or thumbs-down verdict on the entire program.
3. Pay attention to needs and timing constraints of the end users of evaluation results.
4. Invest in capacity building and infrastructure to disseminate, explain, translate, communicate, and use program evaluation results.
5. Revise federal rules to maximize the use and impact of program evaluations.
6. Incentivize the use of cost-effective and evidence-based educational evaluations and educational programs.
7. Establish a Chief Evaluation Officer at the U.S. Department of Education.
8. Fund the right kinds of evaluations.

Each recommendation on its own would be helpful, but taken together they constitute a comprehensive and robust program evaluation system. By investing just a penny on the dollar, the Department of Education can build a robust and rigorous system of program evaluation focused on, and able to improve education outcomes for, all our nation's children.

INTRODUCTION

At a time when our nation is facing enormous social and economic shifts—as well as budget constraints at all levels of government, an increasingly competitive workforce, and growing demand for government services—young people, their families and communities are especially in need of quality, evidence-based educational opportunities. Too often, however, local, state, and federal policymakers have to rely on limited, incomplete, or biased information when deciding how to invest taxpayer resources. In education less than one penny of every dollar spent by the federal government is used to develop the most basic evidence that the rest of that dollar is spent wisely.

The situation raises several important questions. How can we get better results with existing funding? How can we figure out how to stop spending taxpayer dollars on programs that aren't getting results? How do we shift funds to programs that do work? The answer: We get government to play Moneyball.

Today, we know more about what works for young people, their families and communities than ever before, but too often, government doesn't use that information to improve its efforts or to make good spending decisions. In order to have an impact on long-term challenges, more must be done to drive resources toward high-impact solutions that get results. Over the last several years, local, state, and federal officials, along with many in the non-profit and private sectors, have taken critical first steps to ensure that taxpayer dollars are invested in solutions that use evidence and data to get better results. But more must be done and there is no better place to start than education.

There is widespread bipartisan and cross-sector agreement that the public education system in the United States needs to improve *continually* in order for the nation to fully meet the challenges and partake in the opportunities of the 21st century. It is also clear that in these challenging fiscal times, this improvement will need to happen by using existing resources more wisely. For this to occur, it will be necessary to accelerate the transition of education from a craft-based to an evidence-based enterprise.

What is Moneyball for Government?

In a now-famous paradigm change, Billy Beane, general manager of the Oakland A's, ignored scouts and baseball tradition and used statistics, data and hard analysis to build championship contending teams despite limited budgets. More widely, business and social science sectors have begun to use predictive analytics, rapid cycle evaluations, and well-designed impact studies in order to determine what works. It's time for the public policy sector to catch up. Moneyball for Government, a project of Results for America, encourages governments at all levels to increase their use of evidence and data when investing limited taxpayer dollars. By playing Moneyball, we can improve outcomes for young people, their families and communities.

This will not only lead to greater impacts but also will enable local, state, and federal policymakers to shift from educational spending to educational investing.

The good news is that the federal government has a powerful, unused tool at its disposal to help achieve this goal—building a knowledge base of what works and using it to shape federal spending. Yes, the federal government currently evaluates some programs, but those efforts are sporadic and insufficient. Instead, the federal government could launch a comprehensive and responsive evaluation effort to demonstrate which federally funded education programs work, for which students, under what conditions—and how existing federal education programs can be improved to maximize their impact and return on investment. The even better news is that this can be achieved by only investing in evaluation one penny of every federal dollar spent on education. In other words, by wisely spending one penny learning what works, the other 99 cents can go much further.

Setting aside 1% of its program funds for rigorous, independent, third-party evaluations would enable the U.S. Department of Education to: evaluate diverse approaches and initiatives with a range of evaluation methodologies; build the evidence base of which interventions are most effective and why, for whom, under what circumstances; and allow funds to be driven toward those initiatives that result in the most success and the highest return on taxpayer investment. These evaluation funds should have the flexibility to be used across programs and be managed responsibly in partnership with independent, third-party institutions to ensure studies are designed and implemented effectively, will yield the most actionable and reliable data possible, and will build a better evidence base, while also driving continuous program improvements. This evaluation provision would also provide members of congress and the administration with reliable information to gauge program effectiveness and to make policy and funding decisions based on impacts that matter.

Currently, enormous amounts of congressional and executive-branch time, money, and effort are spent on designing, authorizing, appropriating, and regulating federal spending on education with very limited guidance on what actually works, under what circumstances, and for which students. Federal policy, though, exerts a strong influence on what happens in America's classrooms, in particular for our nation's disadvantaged students since their schools rely more heavily on federal funding. Hence by shifting just 1% of federal education funds to a rigorous and thoughtful approach to program evaluation, each step of the process by which federal dollars are translated into education programs could be much better informed and driven by evidence of what works.

Investing in evaluations of federally funded programs has strong bipartisan support. President George W. Bush's administration put a priority on improving the performance of federal programs and encouraged more rigorous evaluations to assess their effectiveness. The Obama Administration has built upon this

effort by supporting an increasing number of evidence- and evaluation-based policies and programs over the past several years. Mayors and governors from both parties are also increasingly using data and evidence to steer public dollars toward more effectively addressing the needs in their cities and states.

Yet, despite this bipartisan support for evaluations, a U.S. Government Accountability Office (GAO) survey of all federal program managers in 2013 highlighted the need for program evaluations across federal departments and agencies. This study found that “only 37 percent of program managers reported that an evaluation had been completed within the past 5 years of any program, operation, or project under their management; another 40 percent of managers reported that they did not know if an evaluation had been completed” (GAO 2013a).

Momentum, however, is building for funding a rigorous and comprehensive approach to evaluation of federal programs:

- On March 14, 2013, the U.S. Senate Budget Committee approved by voice vote an amendment to the Senate’s FY14 Budget Resolution offered by U.S. Senators Wyden (D-OR) and Stabenow (D-MI) that encourages the federal government to use “performance data and scientifically rigorous program evaluations” when reforming, consolidating, or eliminating federal programs.
- On June 20, 2013, *The Atlantic* published an article by Peter Orszag, former OMB Director in the Obama Administration, and John Bridgeland, former White House Domestic Policy Council Director in the G.W. Bush Administration, which called for investing 1% of existing federal funds in evaluations.
- On January 19, 2014, President Obama signed the bipartisan FY14 Omnibus Appropriations Act, which allows the U.S. Secretary of Labor to set aside 1% of funds from all major DOL programs for evaluations and which clarifies the U.S. Department of Education’s authority to reserve up to 0.5% of Elementary and Secondary Education Act (ESEA) funds—except Title I funds, Title III funds, and funds for programs that already have an evaluation provision—for evaluations of ESEA programs without respect to the source of those funds.

The following paper builds on this momentum by outlining why and how the U.S. Department of Education should leverage a 1% investment in evaluation—a penny on the dollar—to develop and implement a comprehensive evaluation system across programs and initiatives. It also details how this comprehensive evaluation system could be designed to accelerate the movement of educational practice from a craft-based to evidence-based enterprise and significantly increase the impact and cost-effectiveness of federal investment in education.

WHY WE NEED TO INVEST MORE IN EVALUATIONS OF FEDERALLY FUNDED EDUCATION PROGRAMS

Ask any group of teachers, principals, school superintendents, school board members, or chief state school officers what would help them provide high quality education to their students, and invariably they will say better information about what works, where, for what type of students, and under what conditions.

When educational practitioners and policymakers, however, are asked how they currently learn what works their responses reveal that education is still primarily a craft-based enterprise. In a craft-based system, decisions about pedagogical approach, instructional materials, educational program, or school/district reform strategies are based more on lived experience, direct observation, and word of mouth from trusted sources than on data analysis or evidence of effectiveness. (Nelson et al. 2009, Asen et al. 2012)

This is not limited to the field of education. Decisions in the military, business, health, and sports fields still often remain driven by craft-based knowledge (Manzi 2012). However, it is also the case that all of these other fields have made more progress than education in moving more towards an evidence-based approach. One reason is that each of these fields makes a bigger investment in evaluation to gather evidence of effectiveness and create an infrastructure for development, interpretation, communication, and application of research and evaluation.

THE U.S. DEPARTMENT OF EDUCATION NEEDS TO LEAD THE WAY

To move education from a craft-based to an evidence-driven enterprise, the U.S. Department of Education needs to take the lead and devote 1% of its program dollars to ongoing program evaluation and to help the field utilize the knowledge its evaluations produce. In addition, the Department needs to build the infrastructure and capacity needed for this knowledge to be gathered, communicated, interpreted, and applied. This fits squarely within the priorities of the U.S. Department of Education. In 2011, the Department released promising practices to state leaders about “how to spend education dollars productively” (Department of Education, 2011). To accomplish this goal, the Department strives to target its efforts where they can do the most good. But without evidence and evaluation, such efforts will fall flat.

Current Approach to Program Evaluation

Unfortunately, the U.S. Department of Education, governed by the laws and funding levels approved by Congress, currently spends relatively little time or resources on program evaluation. The Department spends, in total, around \$45 billion a year—outside of mandatory funding streams including Pell Grants and student loans. Over half of this goes to two programs that spend over ten billion a year: Title I grants to states for educating disadvantaged students and Special Education grants to states. Another 20% goes to six other programs which spend between \$500 million and \$2.5 billion dollars per program per year. These include: Title II-Teacher Quality State Grants, Office of Post-Secondary, 21st Century Community Learning Centers, Career and Technical Education grants, English language acquisition, and School Improvement Grants. In total, eight programs spend nearly 75% of the Department’s annual funding outside of Pell grants and student loans.

Among these programs, in fiscal year 2012, only \$3.2 million dollars were specifically set aside for program evaluation as a line item in the Department’s budget (in this case for the Title I evaluation). Additional funds were also set aside for evaluations, though not specified as such in the budget. Further, approximately \$14.6 million of ESEA funds were set aside in FY12 specifically for evaluations under the Secretary’s authority to collect such money.¹ Taken together though, the total amount spent on evaluations in FY12 represented less than one-tenth of a penny on the dollar of federal spending on education.

¹ This information was compiled from a statement by the U.S. Secretary of Education submitted for the record to the U.S. Senate Labor, HHS, Education, and Related Agencies Appropriations Subcommittee on April 17, 2013.

Program evaluation is supported through other parts of the U.S. Department of Education's budget. This includes funds spent on comprehensive centers (\$50 million), regional education labs (\$57 million), special education studies and evaluation (\$11 million) in the Institute of Educational Science (IES) and IES research funding more broadly. There are also provisions attached to the education funds distributed to states and districts, which allow a small percentage of those funds to be directed toward evaluation. Most of these funds, however, primarily support field initiated evaluation work, that is evaluations proposed by grant or contract recipients or tied to a specific initiative. In a diffuse manner, these field evaluation efforts can help build the knowledge base of what works, but provide neither timely nor sufficient insight into how to optimize and increase the impact of the major federal investments in education. Plus, many of the smaller programs fall through the cracks. A 2009 GAO study found that only two-fifths of the Department's teacher quality programs had evaluations completed or underway, while 11 programs had not been evaluated in more than seven years (GAO 2009).

Movement in the Right Direction

Recent actions by the U.S. Department of Education are trying to change the status quo. The Department, for several years, has been laying the groundwork to help move education from a craft-based to an evidence-based enterprise. For example, the Investing in Innovation (i3) program has taken the lead in showing how rigorous third-party evaluations, that meet the Department's What Works Clearinghouse™ standards, can be built into educational programs. In many cases the Department is dedicating up to 20% of program funds to execute these evaluations to identify what works for which students and under what conditions. In addition, over the past several years, IES has built a new line of inquiry aimed at developing partnerships among researchers and state and local education agencies, as well as developing state and local infrastructure to enable effective program and policy evaluation. For example, IES recently added a component on continuous improvement efforts based on shorter-cycle evaluations of planned variations in educational practice. Similarly, in their most recent re-organization, IES required Regional Education Labs to develop partnerships between researchers, practitioners, and state and local policymakers, in order to help state and local education agencies use longitudinal data to understand their educational challenges and improve their educational practice.

At the same time, technical and technological advances in program evaluation (e.g., online surveying, easier access to administrative records, etc.) are enabling evaluations to be more productive and less expensive. The capacity to conduct high quality, rigorous evaluations has also been significantly aided by the large federal investment in building state longitudinal data systems and the decade-plus investment by IES in building researcher knowledge on how to implement and analyze randomized field trials.

All these efforts, as promising as they may be, fall short of the comprehensive and responsive program evaluation system needed to enable federal education funds to be spent more wisely and achieve greater impacts. Such efforts do not ensure that the full range of education programs are evaluated regularly, that data and evidence are used to continuously improve those programs, or that funding decisions take into account evidence of program effectiveness. More can and must be done in order to get more bang for our education buck.

LEVERAGE EVALUATION TO MAXIMIZE IMPACT: A STRATEGIC FUND TO INFORM POLICY AND PRACTICE

The examples above make a clear case that there is under-investment in federal education program evaluation and a less-than-comprehensive approach to the evaluations that are conducted. Thus, using program evaluations to improve federal decision-making and program outcomes will involve more than just increasing the amount for evaluations. In order to make a difference, evaluation of federal education programs needs to be done strategically.

To accomplish this, the U.S. Department of Education should be authorized to set aside 1% of its program funds for evaluations. The Department should pool these funds and invest them in a manner tailored to the rhythm of federal policymaking. This would enable new ideas to be tested before they are heavily funded. It would also allow program design to be rapidly improved, and enable the building of an evidence base to establish standards of practice. All evaluation efforts supported by these funds should focus on learning what works, where, for whom, and under what conditions. The evaluation funds should also be used to build the capacity of state and local educators to help further develop and use the knowledge base built from the evaluation effort.

Many potential users of program evaluations—federal program managers and administrators, Congressional staff, state and local administrators—do not find current program evaluations useful, timely, or sufficient to direct their actions. By and large, the consumers of program evaluation are not demanding more investments in them because they are not happy with the current product. (See Nelson et al. 2009, Tseng 2012, and GAO 2013a.) Thus, any additional funds invested in program evaluations need to be paired with an effort to make evaluations more useful to practitioners. Such efforts might include speeding up the publishing of evaluation findings, releasing interim evaluation findings that can be used for program improvement, producing evaluations in formats that a broad spectrum of consumers find useful, and aligning the cycle of evaluations with the cycle of school budgets and school years. These issues are discussed below, but they are an essential first step to take to ensure that a strategic evaluation fund is successful.

POLICY RECOMMENDATIONS

Fortunately, there are eight clear steps that the U.S. Department of Education can take with such funds to maximize the impact of program evaluations. While each step on its own would be helpful, together they create a comprehensive and robust program evaluation system that would enable the U.S. Department of Education to play Moneyball.

1 Use scientifically rigorous methods to evaluate programs and build a cumulative evidence base.

To many practitioners and policymakers, program evaluation is a complex, sometimes arcane enterprise, which often results in mixed findings that are difficult to interpret or apply. This in turn fuels a cynical belief that you can find a study to support or reject any program or initiative. The result is that research and evaluation are treated as one of many potential sources of evidence to be used to determine program effectiveness, with no greater validity than personal experience or the word of a trusted source (Asen et al. 2012, Finnigan et al. 2012, Nelson et al. 2009). To move beyond this, education evaluations have to use current, scientifically rigorous methods, but also be purposefully designed to create a cumulative evidence base that is usable, persuasive, and definitive.

The 2013 GAO report on federal program evaluations referenced above found that program managers need more than one study to change course, and that the end goal of program evaluations needs to be “accumulating a knowledge base from which to respond to varied questions over time or fast-breaking policy discussions.” In both medicine and business, for example, multiple randomized field trials are usually required to shift practice in the field.

2 Focus program evaluations on finding what works, where, when, how, and under what circumstances, rather than a thumbs-up or thumbs-down verdict on the entire program.

A primary goal of a comprehensive and responsive approach to evaluation should be continuous improvement of programs. Improvement requires shifting funding and energy within the program to efforts that work, have impact, and have a good return on investment. This also means shifting money and attention away from ineffective efforts. In order to support continuous improvement, the nature of education evaluation must become more sophisticated.

Two realities influence how successfully the federal government can shift its evaluation approach. First, major federal education programs are multidimensional. Even within a single program there

are many components and multiple factors that cause variation in program impacts. For example, the School Improvement Grant program, within Title I of ESEA, operates at the state, district, and school level, with dozens of requirements and options carried out at each level. Further, the program's success is dependent on multiple variables (e.g., school leadership, governance, finance, or scope of student challenges). A single thumbs-up or thumbs-down evaluation of SIG overall is impractical and not useful for the vast majority of practitioners and policymakers. However, thoughtful analysis of key program components at the school or district level—how well they work, for whom, and under what conditions—would yield invaluable information.

An example of this type of comprehensive analysis is a recent study of after-school programs. The study found that a subgroup of programs had large favorable impacts, whereas there were no identifiable effects for others. Moreover, the study found that the effective after-school programs shared four characteristics that could, if implemented correctly, help the less-impactful programs improve (Weiss et al. 2013).

The most valuable evaluation information is timely, as well as relevant and useful. Traditionally, the evaluation process has been focused on summative impacts and not program improvement. These types of evaluations require a great deal of time to put out a request for proposals, select an evaluator, design and implement the evaluation, analyze and write up the results, and then receive clearance to publish the findings. Thus, many federal evaluations have served as autopsies of programs whose political clock has run out, rather than a tool for improvement. Instead, initial and interim evaluations and ongoing data analyses are needed and should be used to inform ongoing administration and implementation of the program, thereby increasing impact and return on investment.

A second reality to consider is the way evaluations are interpreted (and often ignored) based on their thumbs-up or thumbs-down design. Focusing program evaluations on a global yes-or-no impact, as has often been the case, can create a dysfunctional dynamic, which leads the evaluation results to have minimal impact. When researchers are primarily charged to find whether a program works or not, this pushes them to evaluate average results. However, behind almost all average results, there are efforts that worked in some circumstances and not in others. If a program is found to be ineffective based on this overarching analysis, proponents and beneficiaries of the program will fight to highlight the evidence from the circumstances in which the program worked to undercut confidence in the global negative finding. This scenario also denies practitioners the opportunity to dig into the results, understand what made the program effective in certain circumstances but not in others, and use those learnings to improve outcomes. A more surgical approach to educational policy would be more beneficial.

3 Pay attention to needs and timing constraints of the end users of evaluation results.

A core recommendation of the 2013 GAO report on how to improve the use of evaluation in federal programs was to “engage stakeholders throughout the evaluation process,” including input into planning evaluations and “disseminating usable messages” (GAO 2013a). A white paper by Vivian Tseng at the William T. Grant Foundation on “The Uses of Research in Policy and Practice” expands upon this point and makes a compelling case that the impact of research and evaluation has been limited, not because it is communicated poorly or does not focus on important problems of practice, but because it does not sufficiently take into account how the users of evaluations “define, acquire, interpret, and use research; and understanding the social ecology that influences these processes.” (Tseng 2012).

Useful and impactful program evaluation of education programs requires the coordination of three very different worlds that operate on three very different time tables—the academic researchers who largely conduct the evaluations, the federal policymakers and program managers who administer and fund federal education programs, and the educators who use and implement the federal funds and programs to improve educational outcomes. In big and small ways these worlds are often misaligned in manners that undercut the usefulness and impact of program evaluation.

One example is the timing of funding for federal programs in relation to the school schedule. School district budgets for the following year are prepared in the fall and winter and finalized and approved by the school board in the spring. Final school budget, staffing, and professional development plans are all influenced by the district budget and are made in the late spring and early summer. Thus, to effectively impact educational efforts, changes in federal education programs based on evaluation findings need to be aligned with the school planning and implementation cycle.

In practice, however, federal education programs are driven by the federal budget cycle, which operates on a different calendar and more often than not is misaligned with the school planning and implementation cycle. As a result, program improvements based on evaluation findings, however beneficial, may be viewed by educators as problematic. This negative reaction to the proposed changes stems from the disruptive timing of the changes, not resistance to improving practices based on evaluation results. Fortunately, the federal government can improve the situation.

4 Invest in capacity building and infrastructure to disseminate, explain, translate, communicate and use program evaluation results.

Central to increasing the use and impact of program evaluation is enabling researchers, policymakers, and practitioners to not only understand each other, but also to convey insights from one domain to the other. This is no small task. Each group uses its own specialized terminology. When a researcher says “effect size,” a policymaker says “continuing resolution,” and a school principal says my “roster chair,” they might as well be speaking to each other in foreign languages.

Yet, these are all key concepts that inform the use of evaluation results. In theory, the research and evaluation offices within state and local education agencies should be poised to provide these important translation services. However, in practice, these units, many of which were historically formed as a result of the program evaluation movement of the 1970s, have been hit by years of budget cuts and are now often understaffed and that staff is undertrained. In a recent survey of local and state research and evaluation units, over half reported they needed additional training to understand and use two of the central statistical techniques of evidence-based research and evaluation methodology, hierarchical linear modeling and item response theory. Thus, it will be necessary to invest in building the capacity and infrastructure needed at both the state and local level for key information learned in program evaluations to flow easily and accurately to its end users through the use of common, mutually understood terminology.

Building capacity is not the sole, or even primary, responsibility of the federal government. States and school districts must prioritize education research and evaluation themselves. That said, federal investment could help build the backbone and infrastructure that is needed. The vast majority of federal education funding is passed onto state and local education agencies. Moreover, these state and local education agencies oversee the schools that will be the primary settings for evaluations of federal education programs. They are also the entities that often must change behavior or practice as the result of evaluation findings for evaluations to have impact. As a result, it is essential that they have the capacity and infrastructure necessary not only to understand and use evaluation results but also to actively participate in them, inform their design, and carry out supportive evaluation and research activities.

As noted earlier, many state and local research and evaluation units are not sufficiently staffed or trained to explain to others in the network either the design or the results of evaluations using current scientific methods. Thus some of the 1% set aside for federal program evaluations needs to be used to build this capacity, through online and face-to-face training opportunities for local and

state research and evaluation units. IES has some experience in building the technical capacity of researchers, which can be expanded.

Another wise investment of federal funds at the state level would be to build upon the Harvard Strategic Data Analysis Fellows program and provide each chief state school officer with their own full-time strategic data analyst that is part of their executive staff. The function of these analysts could be to both interpret and contextualize the findings of evaluations of federal education programs for chief state school officers and to use the state's statewide longitudinal data system (SLDS) to replicate and extend findings using state level data.

SLDS in themselves represent a major federal investment in state data systems. But they will only be impactful if they are used, and the federal SLDS program should place a priority on the use of data systems to drive improvement. Placing a strategic data analyst at the elbow of the chief state school officer will increase this likelihood dramatically. Finally, by having a strategic data analyst in each state agency, states would have the capacity, on a voluntary basis, to use their SLDS to participate in evaluations conducted at the regional level or among a group of states, if the region or states so desired.

5 Revise federal rules to maximize the use and impact of program evaluations.

Evaluations of federal education programs are governed by a set of federal rules, which go by the acronym of EDGAR, or Education Department General Administrative Regulations. Within EDGAR, there are old rules that stand in the way of carrying out, interpreting, or using impactful evaluations, and that need to be revised. First and foremost is a rule put in place in the 1980s, which states that the U.S. Secretary of Education can grant funds for no longer than 60 months. Evaluations of complex federal programs using up-to-date scientific techniques, however, can easily push past a five-year time limit.

This is especially true when one is studying impacts at the school level on students over multiple years. It would seem essential to ESEA funding, for example, to know what strategies are most effective, under what conditions, for which students in enabling all or most students to read at grade level by third grade or graduate from high school prepared for college and career. In both cases, it is clear that high-needs students will need multi-year interventions. For some students, reading on grade level by 3rd grade will require a pre-K to 3rd grade intervention (a five-year effort). Graduating students from high school will require interventions and continuous supports from 6th to 12th grade (a seven-year effort). When planning and analysis time are added some evaluations will require ten, not five, years of sustained effort.

Congress and the Department have taken some limited action to address this problem. In August 2013, new EDGAR regulations were adopted to allow for 72 months of data collection. The extra time can be used to “fund data collection periods after the end of the substantive work of a project so that project outcomes could be assessed using data from the entire project period” (Department of Education 2013). Further, in January 2014, Congress used an appropriations bill to give the Secretary greater flexibility in grant management. Now the Secretary may renew grants for an additional year, for a total of six years, if the grantee is meeting performance targets (U.S. Congress 2014).

These changes represent a step forward. But they are incomplete absent a more comprehensive fix. For example, at times it will be necessary to extend support for the intervention being evaluated as well, particularly in randomized field trials where the school or district is essentially asked to help validate a promising but not fully proven strategy or intervention. Thus, what is ultimately needed is a more general revision, which states that the Secretary, for high priority or initially promising evaluation efforts, has the authority to grant funds beyond the 60-month limit. Funds would cover the costs of the evaluation but also, where necessary, at least part of the cost of the intervention.

6 Incentivize the use of cost-effective and evidence-based educational evaluations and educational programs.

The methods by which program evaluations are conducted—as well as the ways practitioners and policymakers employ them—need to change in order to create more useful and impactful program evaluations. Traditional program evaluation is time and labor intensive, often involving multiple phases, including evaluation design, measurement development, participant recruitment, observation protocol and survey development, data collection, and data analysis. In addition, randomized designs often involve a substantial education and persuasion effort to gather the layers of approval and buy-in needed to randomly assign interventions to a pool of students, teachers, and/or schools.

Incentives need to be built into the evaluation process to expedite the time and person power high-quality evaluations require. On the researcher side, this could include bonus points for developing or employing lower-cost designs or faster means of conducting evaluations (e.g., using administrative datasets and existing survey instruments). On the participant side, it could involve providing state and local educational agencies with incentives built into the federal funding they receive to routinely participate in randomized field trials and reserve space for a limited number of additional questions in surveys they already conduct to be used for federal education evaluations.

Finally, cost-effectiveness needs to be built in as a routine variable that is examined as part of

federal program evaluations. This is essential in our current environment of limited funding in the face of pressing, unmet educational challenges. The cost of a given program or intervention needs to be compared to alternative programs or interventions, as well as to the status quo. In addition both direct and indirect benefits need to be considered, giving policymakers a sense of their return on investment. For example, two programs might have equal impact on raising student achievement, with one being more expensive. However, if the more expensive program also cut down on special education placements and grade retention (both expensive outcomes) it would be the more cost-effective of the two programs.

7 Establish a Chief Evaluation Officer at the U.S. Department of Education.

In order to carry out an enhanced evaluation program, it will be necessary to have a senior staff member within the U.S. Department of Education who has the resources and authority needed to carry it out. Hence the Department should designate a Chief Evaluation Officer whose primary duty would be the development, management, coordination and implementation of the agency's program evaluation activities to build evidence of program effectiveness.

The Institute of Education Sciences currently serves as the chief evaluation arm of the U.S. Department of Education. As noted above, however, the current approach to evaluation of federally funded programs is limited in scope and funding, and evaluation functions and resources are currently shared between the Department and IES. This leaves some confusion as to who is the person chiefly responsible for evaluating federally funded education programs and how they will be carried out each year.

As the Department considers how to improve its work in this area, it could benefit from looking at the model established by the U.S. Department of Labor. The FY14 omnibus appropriations act, for example, authorized the U.S. Secretary of Labor to set aside up to 1% of its program funds for evaluations to be overseen by its Chief Evaluation Officer. The Department of Education could benefit from greater clarity about who its chief evaluator is, possibly along the lines of actions taken by the U.S. Department of Labor.

Further, the Department should keep a few recommendations in mind.

First, the Chief Evaluation Officer should work closely with Department staff to design, fund, monitor and carry out program evaluations to improve the effectiveness and efficiency of the Department and increase its performance-based initiatives. The Chief Evaluation Officer should ensure that such evaluations meet high quality evaluation standards and are independent, credible and impartial.

Second, the Chief Evaluation Officer should also be responsible for providing information on how evaluation findings helped the Department make informed budget, policy, and management decisions and ensure that the evaluation findings are publicly available and accessible on a dedicated part of the Department's and agency's website. This work is critical to ensuring that evaluation findings and evidence-based approaches are utilized not just by IES but also across the Department in order to improve a broader range of federal programs.

Finally, the Chief Evaluation Officer should engage stakeholders, including the Congress, throughout the evaluation process. He or she could accomplish this goal by organizing meetings, workshops and other discussions to establish a dialogue on evaluation results, including researching ways to ensure effective and appropriate communication of evaluation findings, lessons, conclusions and recommendations. This would also ensure findings are disseminated to appropriate parties so that best practices can be replicated.

8 Fund the right kinds of evaluations.

A comprehensive approach, as outlined above, would vastly improve the impact of evaluations of federal education programs. As the U.S. Department of Education moves forward in this way, it should also pay attention to the specific types of evaluations it funds. At least three types of evaluations will be required.

1 New Policy and Program Validation/Probe Studies

First, both Congress and the Administration need to initiate policy and program probe studies. These are akin to the validation studies in the tiered-evidence framework adopted by the Investing in Innovation (i3) program. These studies would enable the testing and refining of promising new strategies or programs prior to substantial federal investment. For example, there is growing interest in blending learning strategies, which combine computer-based and classroom instruction in new ways. A compelling, even exciting, case can be made for the logic behind these efforts. There are examples in place and some potentially promising early results.

However, there is also a long history of technology under-delivering and failing to solve instructional challenges. Thus, there is the potential for time, effort, and resources to be diverted toward untested approaches that may not result in better student outcomes. A series of well-planned randomized controlled trials could, however, within a two-year period, provide initial insight into the conditions needed for blended learning to have maximum

impact on student achievement. With this knowledge in hand, existing federal programs and new initiatives could be shaped to help spread what works. This two-year analysis may also find that blended learning, in its current structure, is not a good federal investment. In that case, federal administrators and congressional appropriators could avoid making major commitments to an effort with low odds of achieving its intended impact.

2 Improvement Studies to Inform Major Program Design

New federal education programs are typically launched as the result of a new administration's priorities and/or Congressional reauthorization of existing programs. In either case, there is a desire to bring the program to a large enough scale to have meaningful impact during the administration's tenure. Recent examples include Reading First in the Bush Administration and School Improvement Grants in the Obama Administration. In both cases, there is need for rapid evaluation that can help guide program regulation, design, and implementation to enhance its initial impact. As a result, speed and timing in evaluation design and implementation matters.

When a new program is launched, an evaluation needs to provide actionable guidance within two years in order to influence how the program is operated. This means in the very year that a program is set up in schools, its initial evaluation should be planned and executed simultaneously. Further, the evaluation needs to be designed so that learnings in the initial years can inform program design, administration, and implementation in the following years. This work can build on the growing number of examples of how multidimensional evaluation efforts can rapidly inform program design, including the continuous improvement work of the Carnegie Foundation for the Advancement of Teaching and the Measures of Effective Teaching study of teaching funded by the Gates Foundation.

3 Comprehensive Evaluation of the Major Federal Education Programs to Establish Standards of Practice to Inform Re-authorization

As noted earlier in this paper, currently eight federal programs spend \$500 million or more a year. Together, these programs consume nearly 75% of the U.S. Department of Education's discretionary funds. Yet, each of these programs is authorized by different pieces of legislation—ESEA, IDEA, HEA, WIA—meaning they are all on different reauthorization schedules.

For each major program, a comprehensive evaluation initiative should be launched that

is keyed to its reauthorization schedule. This effort should have two goals. First, it should inform budget decisions on an annual basis by building an ongoing body of knowledge that identifies the most effective components, strategies, or efforts within the larger program. Such an ongoing evaluation can ensure that resources are used to incentivize practitioners who are supported by the program funds to adopt evidence-based practices in place of less effective ones.

Second, a comprehensive evaluation initiative should build a cumulative knowledge base that more definitively establishes, within the program, evidence-based standards of practice. In other words, conduct multiple and progressive studies around key questions of practice until it is possible to identify efforts that can be viewed with a high degree of certainty to consistently yield positive outcomes for students and educators. These evidence-based standards of practice could then be incorporated into the reauthorization of the program.

Reauthorization then can become the means by which the major federal programs are progressively made evidence-based. For example, extended learning time is a central component of 21st Century Community Learning Centers program, a billion-dollar program authorized by the Elementary and Secondary Education Act. It is also a key component of many School Improvement Grants, and is an allowable use of funds under Title I. Currently, there is no comprehensive evaluation of the impact of extended learning time in federal education programs. As a result, large sums of federal money are spent each year based on craft-based knowledge supplemented by small-scale, limited-focus, one-time studies.

A better approach would be to conduct a series of linked evaluations over a five-to-seven year period (the typical reauthorization cycle). Some would test different theories of action, and others would build off the findings of initial studies, until it is possible to answer very practical and impactful questions of practice with a large degree of certainty. For example, is it more impactful to extend the school day, the school week, or the school year? Or are all types of additional time equally beneficial? What is the best mix of academic instruction, socio-emotional skill building, and cognitively rich enrichment activities to be employed during extended learning time for different outcome goals, such as raising student achievement, reducing dropout rates, or increasing college completion?

CONCLUSION

In the 21st century, a nation's prosperity is being shaped by the human capital of its citizens. There is little work for a high school dropout, and it is difficult to support a family with only a high school diploma. Therefore, in order for our nation to succeed, our public education system must educate all its students so that they graduate high school prepared to succeed in post-secondary schooling or the workforce. This is a new educational mission, different than the one the great American school system of the 20th century was designed to accomplish. As a result, the public education system has to improve but also change in ways that will bring it into uncharted terrain.

More than ever, we need to build our understanding of which educational approaches, strategies, and interventions, work for which students and under what circumstances. We can only learn this if we formally evaluate efforts and adopt an evidence-based approach to federal investment in education. For this to occur, we need to first build the evidence base, by becoming serious about program evaluation.

By taking just 1% of the money government investments in educational programs and spending it on comprehensive evaluation, it is possible to significantly improve the federal government's ability to invest its dollars wisely and with strong results. Given that the federal investment in education is targeted at providing additional supports to disadvantaged students, such an effort could also enable public education to play an even greater role in providing equal opportunity to all of the nation's children. This is especially critical in an era of great need, declining social mobility, and limited federal resources.

The time is right for the U.S. Department of Education to take the lead and invest 1% of its program funds in a comprehensive and responsive evaluation system. This paper argues that this system should take into consideration what we know is needed for evaluations to have impact and to meet the needs and rhythms of federal policymaking. By funding probe studies of new ideas, rapid improvement efforts when major new educational programs are launched, and comprehensive multiple-study evaluations of existing major programs tied to reauthorization and aimed at establishing standards of practice, the penny spent on evaluations can be leveraged to make the other 99 much more impactful and meaningful to the nation's success. To meet the educational challenges of the 21st century we will have to become both smarter and wiser.

REFERENCES

- Asen, R., D. Gurke, P. Connors, R. Solomon, and E. Gumm. 2012. "Research, evidence, and school-board deliberations: Lessons from three Wisconsin school districts." *Educational Policy* 26.
- Finnigan, K., A. Daly, and J. Che. 2012. "The acquisition and use of evidence district wide." Paper presented at AERA Vancouver April, 2012.
- Manzi, J. 2012. *Uncontrolled: The Surprising Payoff to Trial and Error in Business, Politics, and Society*. New York: Basic Books.
- Nelson, S., J. Leffler, and R. Hansen. 2009. *Towards a Research Agenda for Understanding and Improving the Use of Research Evidence*. Portland, OR: Northwest Regional Educational Laboratory (NWREL).
- Results for America (RFA). 2014. "RFA Investing in What Works Index, May 2014." Accessed 2 July, 2014. <http://moneyballforgov.com/content/uploads/2014/04/RFA-Investing-in-What-Works-Federal-Index-May-2014.pdf>.
- Tseng, V. 2012. "The Uses of Research in Policy and Practice." Society for Research in Child Development (SRCD) Social Policy Report 26(2).
- U.S. Congress. 2014. "Consolidated Appropriations Act, 2014, HR 3547." Accessed 2 July, 2014. <http://www.gpo.gov/fdsys/pkg/BILLS-113hr3547enr/pdf/BILLS-113hr3547enr.pdf>.
- U.S. Department of Education. 2011. "Department of Education Provides Promising Practices for Productivity, Flexibility." Accessed 8 July, 2014. <http://www.ed.gov/news/press-releases/department-education-provides-promising-practices-productivity-flexibility>.
- U.S. Department of Education. 2013. "Federal Register 78(56), 13 August, 2013." Accessed 2 July, 2014. <http://www.gpo.gov/fdsys/pkg/FR-2013-08-13/pdf/2013-19390.pdf>.
- U.S. Government Accountability Office (GAO). 2009. *Sustained Coordination among Key Federal Education Programs Could Enhance State Efforts to Improve Teacher Quality*. Washington, DC: GAO.
- U.S. Government Accountability Office (GAO). 2013a. *Strategies to Facilitate Agencies' Use of Evaluation in Program Management and Policy Making*. Washington, DC: GAO.
- U.S. Government Accountability Office (GAO). 2013b. *Education Research: Further Improvements Needed to Ensure Relevance and Assess Dissemination Efforts*. Washington, DC: GAO.
- Weiss, M.J., H.S. Bloom, T. Brock. 2013. "A Conceptual Framework for Studying the Sources of Variation in Program Effects." MDRC Working Papers on Research Methodology. New York: MDRC.

Acknowledgement

Results for America, an initiative of America Achieves, wants to thank Bloomberg Philanthropies for its generous support for the production of this report.