Center for American Progress

doing what works

The Secret to Programs that Work

New tools for program design and evaluation

Jitinder Kohli, William D. Eggers, and John Griffith   February 2011

Doing What Works Advisory Board

Center for American Progress

The Secret to Programs that Work

New tools for program design and evaluation

Jitinder Kohli, William D. Eggers, and John Griffith   February 2011

doing what works

CAP's Doing What Works project promotes government reform to efficiently allocate scarce resources and achieve greater results for the American people. This project specifically has three key objectives:

- Eliminating or redesigning misguided spending programs and tax expenditures, focused on priority areas such as health care, energy, and education

- Boosting government productivity by streamlining management and strengthening operations in the areas of human resources, information technology, and procurement

- Building a foundation for smarter decision-making by enhancing transparency and performance measurement and evaluation

This paper is one in a series of reports examining government accountability and efficiency.

Contents

Introduction and summary

Less than one-third of Americans have confidence that the federal government can solve problems, according to a recent Center for American Progress survey. The sentiment may be worse at the state level. The Pew Center on the States recently found that less than 20 percent of respondents in California, Illinois, and New York trusted their state governments.

What accounts for such widespread frustration?

It's not just the economy. As Vice President Joe Biden has said, there's a feeling across America that "Washington, right now, is broken."

And in some ways, it is.

Washington for years has been shooting at big targets and continues to miss the mark. After spending more than $1 billion, the government last month scrapped a troubled "virtual border" plan plagued for years by cost overruns and delays. In preparation for the 2010 Census, the Commerce Department spent two years developing handheld computers to replace millions of costly forms and maps used by field workers. The initiative failed and workers reverted to pen and paper.

At the root of such failures is faulty design. The way public policy is designed today results in programs that sound good in hearings but don't work in the real world. This paper diagnoses common design flaws, and proposes a kind of advance-warning system to help policymakers distinguish between programs with a high chance of success from those likely to run into problems down the line.

Five common design flaws in government programs

After consulting about 200 government experts over six months, we discovered a handful of common problems that can doom government programs before they begin.

The wrong approach

Decision makers often forget to ask the most basic question when considering new programs: Do we really need to do this?

Insufficient evidence

Government programs too often sprout from little more than a policymaker's hunch, with scant evidence they'll actually work.

Poor implementation planning

Program advocates often get so caught up in the political process that they skip over crucial implementation issues in the design phase, leading to cost overruns and timeline delays.

Misunderstood incentives

Programs created without a precise understanding of the incentives embedded in them are vulnerable to deception or "gaming."

Insufficient performance assessment and refinement

The best businesses are committed to constantly monitoring and improving the performance of their products. But not enough government programs are designed to report whether they're actually working.

A checklist-inspired solution: Design for Success

In his book, *The Checklist Manifesto: How to Get Things Right*, Atul Gawande revealed the power of simple checklists to prevent systemic failure in a variety of contexts, from operating rooms to investment companies.

Gawande's thesis was a starting point for this project. Could a checklist-type system be used to predict a government program's likelihood of success? Might it prevent the all-too-common design flaws that lead to implementation crises?

In this spirit, we sought to define the characteristics of a program that was likely to succeed and then ask proponents to go through a checklist-type process early on.

An effective checklist has two components: A list of well-designed questions, and a process for how and when to ask them. We narrowed our focus to five broad areas, covering questions that can help policymakers avoid common design flaws:

- **Approach:** Is this the right approach to address the problem?
- **Evidence of effectiveness:** Has the program been successfully implemented elsewhere? Has there been a rigorous evaluation of its impact?
- **Incentives:** Does the program design minimize the risks of cheating the system?
- **Implementation:** How does the agency responsible for administering the program plan to secure the necessary staff, skill base, and technology infrastructure? Are the plans and timelines reasonable?
- **Monitoring and rethinking:** Are there clear indicators that define success? Is there a plan for collecting timely and accurate data to monitor performance?

These questions underpin a series of questionnaires and checklists that probe the five common design flaws we identified.

Of course, our "Design for Success" tools themselves are only valuable if used the right way. So we also dedicated considerable time to working with experts on how to best fit these tools into the program-creation process. Here's the process in a nutshell:

STEP 1: The new program proponent uses a "program checklist" to design an initiative.

STEP 2: The proponent completes a "program details" questionnaire that probes the five key components of a successful program described above.

STEP 3: A neutral party, such as an interagency panel or legislative committee, completes a "program assessment" questionnaire to evaluate the likely success of the new program.

STEP 4: Decision makers use the information on both questionnaires to guide their scrutiny of the program.

Reviewing What Works: Tools for existing programs

Having established a process to predict the likelihood of a new program's success, we next adapted these tools and procedures to an equally important task: evaluating the effectiveness of existing programs.

At a time of looming budget cuts, Washington urgently needs a better way to distinguish the most effective programs from those in need of reform—else it risks slashing good programs simply because they have less political support.

Our tools for evaluating existing programs build on two major recent government performance milestones: The Obama administration's recently adopted 128 High Priority Performance Goals, and the Government Performance and Results Modernization Act, signed by the president in January. The law requires the executive branch to adopt cross-cutting outcome goals and to report regularly on progress toward achieving those goals.

The "Reviewing What Works" process evaluates programs across a policy area against these goals, using interagency panels as arbiters of effectiveness. Again, the process in a nutshell:

STEP 1: The government forms interagency panels by policy area.

STEP 2: These panels define common goals and list programs that contribute to these objectives on a "policy strategy" questionnaire.

STEP 3: Program managers complete a "program effectiveness" questionnaire for each initiative.

STEP 4: The interagency panels complete "program evaluation" questionnaires to determine the effectiveness of individual programs.

STEP 5: The questionnaires inform decisions about which programs to expand or reform.

In the existing programs context, the key questions revolve around five key concerns:

- **Impact:** What impact does the program have on the goals across government in the particular policy area?
- **Collaboration:** Does the program coordinate with other programs to maximize collective impact and minimize duplication?
- **Benchmarking:** What is the relative effectiveness and cost of the program?
- **Operational excellence:** Is the program well run? Have there been delays or cost overruns?
- **Adaptability:** Has the program sought to learn from experience? Has it improved in response?

A time to act

More than 80 percent of Americans think the federal budget process should be reformed so that spending decisions are based on what works, according to a 2010 Center for American Progress survey.

This demands more prominent consideration during the design phase about whether a new program is likely to work. Equally, we need a system that scrutinizes existing programs for effectiveness, not merely political attractiveness.

We can no longer defer proper consideration of which programs are most and least effective. The time to act is now. We believe this report shows a way forward.

Background

More than 400,000 temporary census workers armed with paper and pencils swarmed the country collecting data for the 2000 Census.[1] The effort cost more than twice the previous decennial count.

Census Bureau officials devised a seemingly simple solution to the soaring expenses: Replace millions of costly forms and maps with handheld computers for the next census. The potential payoffs were massive. Data collectors could instantly transmit data to a shared database, update residency information from anywhere in the country, and more easily navigate routes with an internal GPS system.

That was the plan, anyway.

Two years after awarding a $600 million contract to develop the handhelds and related software, Commerce Secretary Carlos Gutierrez announced in April 2008 the bureau would scale back its use of the devices. The agency requested $2 billion to $3 billion in additional funding to help data collectors transition back to paper.[2]

It was a predictable blunder, IT expert Robert Charette told Government Executive magazine: "A blunder is when you don't do the things you know you should, many times because of hubris, and things go off the rails, predictably."[3]

The aborted transition to handheld devices was not a failure of technology. It was a failure of design.

The No Child Left Behind Act of 2001, the 1986 immigration reform, and the price and wage controls of the early 1970s all suffered from similar problems of poor design. They were launched with high hopes of social and economic progress, and all had results so disappointing that politicians soon found themselves back at the drawing board.

The past few years have been no different. The 2009 Cash for Clunkers program drastically underestimated potential demand, leading to too little money allocated and too few workers to process applications. Likewise, SBInet, the planned $8 billion "virtual border" fence, was rushed into implementation without a well-thought-out plan and was "plagued with cost overruns and missed deadlines," according to Janet Napolitano, secretary of the Department of Homeland Security. The high-tech fence was recently abandoned by the Obama administration.

Similar problems have plagued complex initiatives at the state level. One of the biggest policy fiascos in recent times was California's electricity deregulation. Democratic and Republican lawmakers in 1996 cooperated on a major redesign of the state's electricity markets. The reforms were intended to introduce competition, spur innovation, and lower the cost of electricity. They were a spectacular failure. The new law caused soaring prices, rolling blackouts, and the recall of Gov. Gray Davis. A government reform initiative launched with high hopes had turned into a disaster.

At the root of each of these failures was faulty design. Just as architectural designs can look good on paper but falter in the real world, poorly designed programs are often a blueprint for disaster.

One consequence of these failures: Americans' confidence in our government has fallen to an all time low. Less than one-third of Americans trust the federal government to solve problems, according to a recent Center for American Progress survey.[4] The news is hardly better at the state level. A recent Pew Center on the States survey found that fewer than 20 percent of respondents in California, Illinois, and New York trusted their state governments.[5]

It's not only voters who believe we have a crisis. So do the government's most senior civil servants. Sixty percent of the federal government's Senior Executive Service said in a 2008 survey that Washington today is less capable of executing large projects than it was 30 years ago.[6]

Time for a change

We're at the dawn of an era of fiscal austerity. Funding for new initiatives will become increasingly scarce—and there will be little margin for error. Couple this with zero tolerance on the part of the American public for the kind of

large-scale failures we've experienced in recent decades, and it becomes clear that Washington needs an advance-warning system to sort programs with a high chance of success from those likely to run into problems down the line.

Washington needs to ask basic questions of every proposed program: Is this the right way to solve the problem? How do you know it will work? How will you implement it?

> "On the Hill, we could really do with a checklist for reviewing new legislation. That would help us make sure that we're asking the right questions of new ideas."
>
> – *Congressional staffer*

That's what this paper is about. After working with some 200 experts, we have developed tools and processes to ensure that these questions are asked and answered early on, both by those responsible for designing new programs and those with the power to decide which programs to back.

While analysis of the likely success of a new program idea will never supplant the need to look at how a new idea plays politically, looking at these issues alongside political considerations will increase the chances that good ideas turn into successful programs.

Of course, improving the design of new programs is not enough. The government also needs a better way for reviewing whether existing programs are working, experts told us time and again. As with new programs, basic questions must be asked of existing ones: What impact has the program had? Does it duplicate other programs? Is it cost-effective? Is it run well? Has it improved over time? We urgently need a formal process to ask these questions and distinguish effective programs from those in need of reform.

To that end, this paper also proposes adapting our tools for new programs to existing programs. Description of our existing program tools follows the presentation of new program tools. Many of the experts with whom we consulted for this project told us that the fiscal outlook meant these tools would be even more important in the coming years.

If we can put a man on the moon…

This report builds on the 2009 book, *If We Can Put a Man on the Moon: Getting Big Things Done in Government,* in which Bill Eggers and John O'Leary examine how governments tackle their biggest challenges. The goal of the book was to answer a seemingly simple question: Why do some big initiatives succeed and others fail?

The authors studied more than 75 major public initiatives since World War II and found that nearly all followed a predictable path from idea to results. From the so-called War on Poverty to the war on inflation, from fixing inner-city schools to putting a man on the moon, the successful public policy journey always includes the same basic elements: A good idea, a well-designed piece of legislation, political support, and strong implementation. (See figure below)

Like links in a chain, every step in the process must work. Failure can occur at any point in the journey. Nixon's wage and price controls failed because they were based on a faulty understanding of infla-tion. California electricity deregulation failed because of bad design. Iraq reconstruction suffered from poor implementation.

The root cause of program failures, albeit a hidden one, is a hitch in the process illustrated below. All the successful initiatives examined by Eggers and O'Leary got the process right.

Proponents took the time to listen to opposing viewpoints and often incorporated critiques into the program's design. Lawmakers were aware they were crafting a design that needed to work in the real world. Sponsors allowed for thoughtful debate. Once the bill was passed, the political champion took the possibility of failure seriously and recruited a strong manager to lead the implementation.

More information on the book, as well as an interactive map of the Journey to Success, is available at www.journeytosuccessmap.com.

The journey to success

Idea	Design	Stargate*	Implementation	Results

Reevaluation
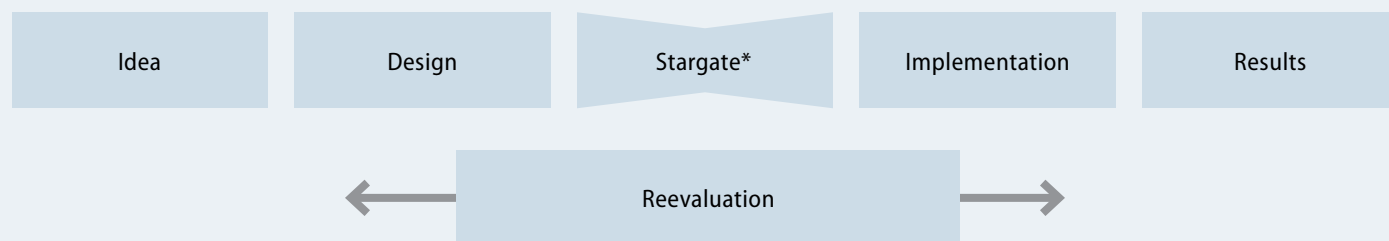
*Stargate: In their book, Eggers and O'Leary define "Stargate" as, "The moment of democratic commitment because it instantly takes the process from the political universe to the bureaucratic universe… Think of it as the moment when a bill becomes law."

The problem: Why programs fail

Failed government initiatives share a number of common design flaws. While most programs aim to inspire some social change, too often they are drafted, debated, and funded without a clear idea of what proponents hope to achieve for the American people.

And even when the goals are clear, programs are often poorly designed. Program "architects" sometimes take the wrong approach to the problem altogether, or

What's a "program"?

This report is about government "programs." What does that word mean?

People in Washington use the term to mean different things, so we should be clear about our definition. We use the term to refer to a discrete initiative or intervention by government to achieve one or more goals. It might be a grant program to help schools or a new website that makes it easier for people to renew their passports. It could be a new regulatory initiative, such as a food-labeling requirement on caloric content to thwart obesity, or a new licensing requirement for particular pollutants to improve air quality. It could be a tax expenditure, such as a tax credit to encourage retirement savings.

Whether it derives from the government's power to spend, tax, or regulate, a government program is an initiative designed to achieve one or more policy goals.

A "program" as we use the term is focused and defined. So a package of measures, such as those in the health care bill or the new financial regulation overhaul, does not constitute a program, but is rather a group of discrete programs. To apply the approach set out in this report, each program in a package of measures would need to be considered separately.

they design programs without thinking through the incentives they're creating. Programs are launched without good evidence that they will work and without a plan for tracking progress toward achieving their goals. Implementation planning often takes a back seat to political palatability.

In many cases of mega-failures, the policy design process takes place in a vacuum, largely disconnected from the implementation process. In the best private companies, the concept of "designing for execution" is second nature: The design department works closely with the manufacturing team to avoid drawing up something that can't be built. Such cooperation is rare in government. More than two-thirds of senior federal executives say there is often or always a disconnect between people who develop policy and those who implement it.[7]

Five common design flaws for new programs

So what are the common design flaws that can doom government programs? Here are five key design flaws, according to our analysis and the experts with whom we consulted:

The wrong approach

Decision makers too often leave out the most basic questions when considering new programs: Do we really need it? Is the problem already being addressed by another federal program? Are we duplicating efforts?

Consider the Department of Agriculture's High Energy Cost Grant Program, which delivers $15.5 million a year to utility providers in Alaska, Hawaii, and several other remote regions. Subsidizing energy services to isolated areas might be a great reason for a federal initiative—but the grant program is completely unnecessary.

That's because the Rural Utility Service, a division of the agriculture department, already administers a loan program that serves the exact same purpose, only more effectively.[8] All of the areas eligible for high energy cost grants are also eligible for electricity loans. President Barack Obama, like his Republican predecessor, has pushed to terminate the duplicative grant program, arguing that the loan program

is a more inclusive and cost-effective way to achieve the same end.[9] And yet the grant program continues to find funding in every appropriations bill.

Even when a new program is necessary, some approaches are better than others. A grant program might make sense in one scenario, while a loan program works better in another. Sometimes tackling a problem through a spending program isn't as effective as a regulatory or tax approach.[10]

Insufficient evidence

Even the most promising ideas are often flawed, based on a fundamental misunderstanding about how the world works. Yet time and again we see new government programs sprout from little more than policymakers' hunches, with no real evidence they'll actually work. Funding decisions for new initiatives are often based more on political suitability than whether they will actually work.

Fundamental questions are often absent from the decision-making process, such as whether the program has been implemented at a smaller scale and how it fared.

Consider the Department of Homeland Security's multi-billion dollar SBInet, the technology component of the government's Secure Border Initiative. The idea was to erect a "virtual fence" along the nation's southwestern border, equipped with state-of-the-art surveillance cameras, radar systems, sensors, and centralized command systems.

It sounded great, but it didn't take long for problems to arise. Daytime human detection cameras were effective at just half of the required distance. Laser range finders had an effective range of less than two kilometers, far below the required ten kilometers. The date of deployment was pushed back by seven years. The costs more than doubled. From March 2008 to July 2009, federal auditors found more than 1,300 defects in the SBInet system, with new problems being discovered at a faster rate than repairs could be made.[11]

The bottom line: This new technology was being rushed into implementation without being fully tested on the smaller scale.[12] Many of these issues stem from an aggressive political push to get the overall SBI program up and running.[13]

In January 2010, DHS Secretary Janet Napolitano froze funding for an expansion of the struggling project, and two months later diverted $50 million in stimulus funds slated for SBInet to more reliable border security tools.[14] After an investment of over $1 billion, the program was effectively halted, with little to show for it. The administration formally terminated this program in January 2011.[15]

Poor implementation planning

According to our survey of federal senior executives, unrealistic costs and timelines are the biggest culprits leading to the failure of policy initiatives.[16] Proponents are often so wrapped up in the political process that they skip through crucial implementation considerations in the design phase.

That leads to a host of problems. Timelines become unreasonable. Necessary staff and information technology structures are not put in place. There's no concrete strategy for delivering services or mitigating major risks. The level of interest from consumers and other key stakeholders is insufficiently gauged. The program is set up to fall apart.

The ramp-up to the 2010 Census is a portrait of failing to design for implementation. The Census Bureau in 2006 initiated the Field Data Collection Automation program to develop more than 525,000 handheld computers for temporary data collection workers. In all, the bureau estimated it would save $445 million compared to the pencil-and-paper method used in 2000.[17]

Management and implementation problems surfaced soon after the program began. Federal auditors in 2008 warned that the bureau had not "implemented the full set of acquisition management capabilities" necessary to manage the program, such as developing baseline requirements and acquisition plans.[18] It soon became clear that key risks—such as device malfunctions in the field— had not been planned for, and a plan had not been developed for measuring the devices' performance.[19]

The failure to design for implementation foreshadowed a cascade of operational problems: Costs increased, deadlines went unmet, and developers failed to deliver fully operational handhelds. The Census Bureau ultimately decided to

return to pencils and paper, using a fraction of the handhelds originally planned, and only to update address information.[20] According to federal auditors, it was mismanagement and lack of planning—not technological issues— that caused the program's demise.[21]

The wrong incentives

Almost all government programs alter behavioral incentives. A grant program does so directly; a regulatory measure often does so indirectly. When the incentive system embedded in a program is well thought-out, you get the behavior change you want. When you don't understand all incentives created by a program, you leave the initiative vulnerable to deception or "gaming."

New programs are particularly prone to such gaming because designers often fail to imagine all the creative ways the program might be scammed. It's important not to underestimate the ingenuity of everyday people, especially when it comes to exploiting flaws in a potentially lucrative system.

Consider the Immigration Reform and Control Act, which Congress passed in 1986. The law established sanctions against employers who hire unauthorized workers and legalized different categories of undocumented workers. Numerous compromises were required to secure final passage of the law, and one of these compromises resulted in a fatal design flaw.

The law required employers to collect eligibility documents from every new hire. It did not, however, require that the employer verify the identity of the person presenting the documents. This sparked an underground market in counterfeit papers enabling undocumented immigrants to secure employment with false or stolen documents.[22] Employers were rarely exposed to sanctions for hiring undocumented immigrants as long as the papers appeared valid.

 So the mechanism designed to hold employers accountable for hiring authorized workers—sanctions for failure to collect eligibility documentation—did not align with the program's goal of actually preventing the hiring of such workers. It also inspired undocumented immigrants to commit fraud by using fake documents. The failure of this law, along with numerous other flaws in the legislation,[23] has left us with a broken system 25 years later in which undocumented workers are 5 percent of the nation's labor force.[24]

Insufficient performance assessment and refinement

Part of designing a successful program is embedding within it a way to know whether it's actually successful in solving the problem it is designed to address. But all too often program proponents don't think about what data is needed to monitor success. And they are so convinced that their program is bound to succeed that they don't build in a strategy to learn from experience, or shut down the program if it proves unsuccessful.

The best businesses are committed to constantly improving their products but few government programs act that way. As a result, programs that are working poorly become part of the landscape.

The No Child Left Behind Act of 2001 is a recent example of the "performance assessment" problem. The bipartisan and popular law was sold as a way to improve education through tougher accountability standards. The law required states not only to introduce testing but also to make progress toward attaining student proficiency. But each state was left to determine what "proficiency" meant.

The theme of "accountability for results" was clear, but how to measure those results wasn't. States defined their own benchmarks, resulting in massive inconsistencies in implementation and reporting across the country. "[The designers of NCLB] were very, very loose about what the goals were, what the goalposts were, the benchmarks," said Education Secretary Arne Duncan in a 2009 interview. "Fifty states did their own thing. That didn't make sense to me."[25]

This fragmentation has made it difficult to assess the impact of the NCLB law across states, despite a wealth of new data on educational performance. Researchers can study the law's effect in individual states, but even that data can be unreliable at times.[26] Surely the new standards imposed by the NCLB law have done some good, but educational performance in the United States remains far short of NCLB's initial promises.[27]

Why don't government programs avoid these pitfalls?

One particularly troubling issue is that these design flaws are well understood by people who develop and study government programs, and yet the flaws recur. Practically everyone we consulted—agency officials, Hill staffers, state govern-

ment managers, academics, Government Accountability Office watchdogs, and new government recruits—agreed on the main causes of program failure, and agreed that these problems were endemic.

So if it's relatively easy to understand the problems to avoid, why don't decision makers focus on them at the critical points of program design and adoption? Surely it would make sense to question whether cost estimates are accurate before appropriating money, or whether an initiative has been successfully piloted at a small scale before giving it a national green light. If decision makers want programs to succeed, why don't they always ask these questions?

We believe there are two main reasons government so often fails to avoid program design pitfalls.

First, program proponents often strongly *believe* that their idea will succeed. Instead of spending their time testing and honing the idea to ensure that its chances of success are high, they spend their time persuading others to go along with it.

Second, proponents tend to focus on the politics and perception of a new idea, rather than on less glamorous questions of whether the program is likely to work or whether it is ready to be implemented. They focus on which stakeholder group might back the idea, how it will play with the media and voters, and what effect it could have on future political contests. These considerations naturally lead to compromises, and ideas get amended to increase political support. The changes, however, are rarely about making the idea more effective when implemented, but about luring the support of powerful players.

The problem, then, is that our program-making process focuses primarily on politics, and only secondarily on substantial policy questions. Questions of implementability sometimes seem entirely absent from the process.

Why is this? Part of the reason is the short time horizon of public officials. It can take years to determine whether a new idea has worked, and by then the political focus may have moved on. The senior official who championed the initiative has often moved on before her pet project was fully implemented. The congressman who backed the program on its passage through the legislative branch has since had his attention diverted by a multitude of political issues. By the time the success or failure of the program is apparent, its design may well be a distant memory in the minds of the people who designed and debated it.

So how can we overcome this problem? The rest of this paper is devoted to exploring our recommended solutions: A checklist-inspired series of tools that require proponents of new programs to ask the right questions at the right time.

Our "Design for Success" tools are intended to be used before *new* government programs are launched. The next section relates how we developed the tool and describes it in detail.

The section after that presents an adapted version of these tools that can be used to evaluate *existing* programs. We call these the "Reviewing What Works" tools, and outline a process for their use that will help government to distinguish between effective programs and those that need reform.

A checklist-inspired solution: The Design for Success tools

In the 2009 book, *The Checklist Manifesto: How to Get Things Right*, author and physician Atul Gawande reveals the power of simple checklists in preventing systemic failures. For example, a five-point medical checklist implemented in 2001 virtually eradicated infections in patients with central intravenous lines in the intensive care unit at The Johns Hopkins Hospital in Baltimore. A checklist designed by a Southern California investment firm required decision makers to consider 70 common pitfalls of bad investments, such as forgetting to consider the impact of "boom" conditions when reading a company's revenues report. After a year of running down that checklist before every investment decision, Gawande reports the firm's investments were up 160 percent on average, without a single avoidable mistake.

Gawande argues that the checklist model can be expanded to many high-stakes professions and activities: "We don't study failures in teaching, in law, in government programs, in the financial industry, or elsewhere," he writes. "We don't look for the patterns of our recurrent mistakes or devise and refine potential solutions for them. But we could, and that is the ultimate point."[28]

Gawande's thesis was a starting point for this project. Could a checklist-type process be used to gauge a government program's likelihood of success? Could it be implemented in a way that ensured proponents and other officials prevented the all-too-common design flaws that plague government programs?

Our approach

An effective checklist has two components: A list of well designed questions, and a process for how and when to ask them. Sounds easy. But devising a good checklist is not easy. It must capture the complex causes of program failure while being simple enough to be useful and easy to use.

We took a consultative approach to this problem, and over six months enlisted the feedback of around 200 experts, to make sure we were asking the right questions and recommending a sensible process.

Our first step was to convene a brainstorming session with about 60 government performance experts, including program staff from agencies, public policy researchers, and state and local government officials. We asked them to design from scratch a checklist that could be used to gauge a proposed new program's likelihood of success.

The initial session left us with an initial draft of the questions our tool would need to ask. But there were conflicting opinions about how to design the questionnaire itself. So we designed three prototypes, each with a distinct format.

The first included 25 questions, each asking for concise narrative responses. The second design collapsed those questions into 10 broader topics, but asked for longer-form responses. The third included a way to assess the program's likelihood of success in each category.

With these three options on the table, we held seven smaller workshops, asking each expert group which tool worked best, why, and whether we had left any important questions unasked. They suggested we should have a detailed questionnaire for proponents of new programs, and also a shorter one for neutral parties to assess the likely success of a program idea. They also gave us extensive feedback on the drafting of individual questions.

Among participants in these follow-up workshops were budget examiners from Office of Management and Budget, agency staff responsible for designing and monitoring programs, academic experts, and people who implement federal programs at the state level. We also sought input from congressional staffers, oversight experts at the Government Accountability Office, and Presidential Management Fellows, who are promising new managerial recruits to the federal workforce.

After that workshop round we designed near-final drafts of the tools, and circulated them again to around 25 of our experts for a final round of comments—and a final round of revisions. The Design for Success tools we present in this report are the product of this process.

The Design for Success tools

The Design for Success tools (see thumbnail images, and the appendices for the complete prototypes) are a series of questionnaires and checklists that seek to flesh out the five common design flaws described in the first section of this paper. Accordingly, the tools ask respondents to answer questions covering five key components of successful programs:

Approach

The first set of questions is about whether the program under consideration takes the right approach to address the problem. Is it possible to expand or modify an existing program? Are there more cost-effective ways to address this need?

Effectiveness

Next, the respondent is asked to show evidence that a program is likely to successfully address the problem. Has it been successfully implemented elsewhere? Has there been a rigorous evaluation of its impact?

Incentives

As we have seen, program designers often fail to think through the incentives their new programs and policies establish. A key question in this section of the tool is whether the program design minimizes the risks of someone "gaming the system."

Implementation

Architects of new programs often pay too little attention to the nuts and bolts of implementation. This section probes whether designers have thought through operational issues such as administration plans, human resources capacity, and information technology infrastructure.

Monitoring and rethinking

Good programs learn—and improve—as they move forward. In this section we ask whether program designers have defined clear indicators of success, and whether there exists a plan for collecting timely and accurate performance data.

Excerpt from the Design for Success: Program Assessment Tool

6. Why will the program work?
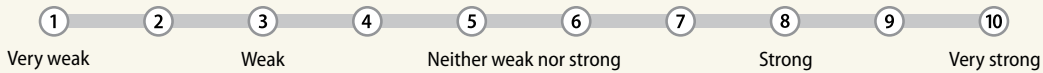
① —— ② —— ③ —— ④ —— ⑤ —— ⑥ —— ⑦ —— ⑧ —— ⑨ —— ⑩

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak evidence

Unreasonable expectation that key stakeholders will alter their ways of working or behavior; similar programs never implemented elsewhere; research on the approach either unavailable or concludes it is ineffective; no plan to test the program before rolling out on the large scale/expanding on the model too quickly

Strong evidence

Reasonable expectation that key stakeholders will alter their ways of working or behavior; professional, independent research indicates approach will work; program expands on what worked on the smaller scale/plan to test before rolling out on the large scale

The Design for Success process

Process matters a great deal in government. For that reason, we spent almost as much time working with our expert community on how to best fit these tools into the program-creation process as we did designing the tools themselves.

Our first hurdle was to recognize and account for the fact that programs emerge from both the legislative and executive branches of government. The processes for the adoption of new programs in each branch are different, and so our tools need to be deployed in different ways.

Executive branch agencies normally devise new ideas for spending programs where they already have legal authority from Congress. Ideas often originate from bureaus in agencies. Central teams coordinate processes to decide which ones the secretary should back. Agencies then put forward their preferred ideas to OMB seeking inclusion in the president's budget. OMB decides which ones to include, and Congress makes final decisions in the appropriations process.

In many cases, however, it's the legislature that decrees programs into being. When this happens, the proponent of the program is typically a lawmaker who conceives of a government initiative, drafts legislation, and guides it through the legislative deliberation and hearing process until it is enacted by both houses.

While we have developed slightly different tools depending on whether the program originates in the executive or legislative branch, the process we recommend for implementing them follows the same basic steps:

Interagency panels

The Design for Success process proposes that interagency panels organized around policy areas serve as "neutral evaluators" of executive branch program design.

The U.S. Interagency Council on Homelessness is a good model for such a panel. This council coordinates the administration's work on homelessness and has developed interagency goals. It is chaired by a cabinet secretary, and has broad cabinet-level membership. Its work also brings together experts from across the administration with those outside government.

"When I was designing a program last year, I wished I had a list of questions like this to ensure that I thought about everything important."

– Federal agency program manager

STEP 1: The proponent uses the "program checklist" to design an initiative
The chief architect of the proposed program and her colleagues use the simple checklist as a guide to good program design. (See Appendix 1)

STEP 2: The proponent completes the "program details" questionnaire
When the program is designed, the proponent fills out a form that probes the five key components of a successful program described above. (See Appendices 2 and 3)

STEP 3: A neutral party completes the "program assessment" questionnaire
The "program details" form is then submitted to a neutral party, which reviews it and completes a corresponding "program assessment" form that captures a judgment about the likely success of a new program. For programs designed in the executive branch, the neutral party might be an interagency panel, such as those already in place to work on multidisciplinary problems like homelessness and obesity. (See box above) For legislative programs, the neutral party could be a legislative committee or independent review body such as the Congressional Budget Office. (See Appendices 2 and 3)

STEP 4: Decision makers use the information on both questionnaires to guide their scrutiny
Lawmakers (and their staff) and executive branch managers review the program detail and program assessment forms. As the proposal advances in legislative hearings or policy-development meetings, decision makers will use this information to highlight areas of weakness and prompt program designers to make changes that maximize the idea's chances of success.

Of course, the journey described above is only part of the process that determines whether a program goes ahead. Policymakers should always consider the political attractiveness and urgency of a new idea. But we believe more effective programs will emerge from a process that also prompts program designers, evaluators, and decision makers to work through questions that are predictive of likely success.

The figure below shows a simplified depiction of how the Design for Success tools would be incorporated into the existing program-creation process.

Design for Success—legislative process

Step 1	Step 2	Step 3	Step 4
Member comes up with an idea and uses the Program Checklist tool to help design program	Member drafts legislation and fills out the Program Details tool	Independent body reviews the Program Details tool and fills out the Program Assessment tool	Decision makers in the committee and the rest of Congress use the information to guide their scrutiny
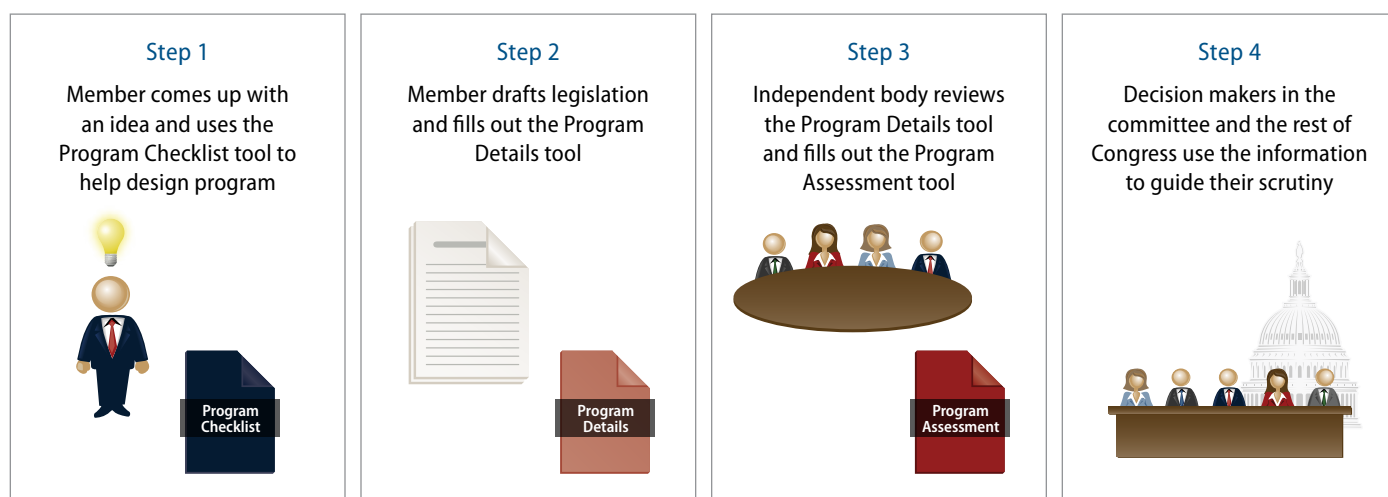
Design for Success—executive process

Step 1	Step 2	Step 3	Step 4
Agency staffer develops idea using Program Checklist tool for assistance	Agency staffer develops a program proposal and fills out the Program Details tool	Interagency panel reviews the proposal and fills out the Program Assessment tool	Decision makers in agency, OMB, and in Congress use the information to guide their scrutiny
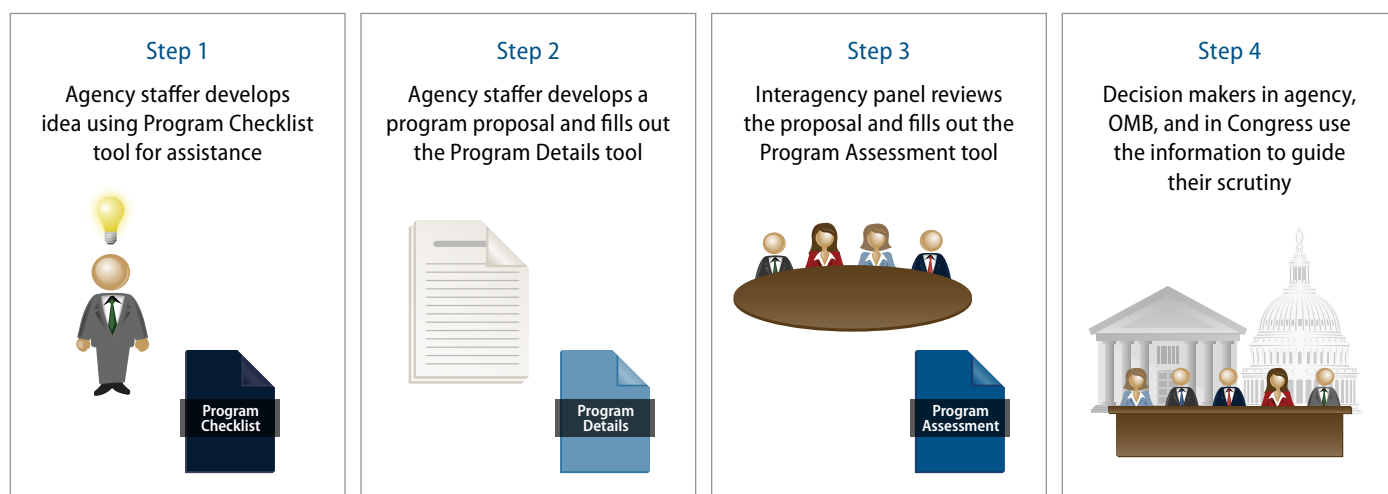
Appendices 2 and 3 include a more detailed discussion of exactly how these tools would be implemented in the legislative and executive branch scenarios.

What about new nonspending programs?

The tools and process presented above are designed for new spending initiatives, so they will not work as designed for nonspending programs such as regulatory or tax initiatives.

Many regulations carry little direct cost for government, and instead impose a cost on the people or businesses affected by the rules. The process of regulatory impact analysis that is already carried out in the executive branch focuses on the costs and benefits of new regulations.[29] In that process agencies do consider many of the questions we think are important for the success of new initiatives, such as whether alternatives have already been considered. But the existing impact analysis omits other important questions, such as whether offices responsible for enforcement are sufficiently staffed.

Appendix 4 of this paper includes an alternative form of our "program details" tool for regulatory proposals, designed to supplement the existing processes within government.

Tax expenditures, commonly known as "tax breaks," are government's way of spending through the tax code, since tax breaks have the same net effect on the budget as direct spending. But Congress often approves new tax breaks without subjecting them to the same scrutiny as direct spending programs.

Appendix 5, therefore, includes a draft tool developed by CAP researchers for evaluating new tax expenditure programs that provide incentives or disincentives for certain activities (such as saving for retirement). It would not be completed where the objective of the provision was entirely tax related. The questionnaire prompts respondents to address performance and implementation areas such as goals, approach, costs, evidence, and incentives to fit these programs.

In the next section, we discuss how to adapt the Design for Success tools to a more immediate problem, given our current fiscal state: How to evaluate the effectiveness of existing programs.

Reviewing What Works:
Tools for existing programs

The best way to attack a problem is to prevent it from occurring, which is why our Design for Success system naturally focuses on the design phase of new government programs—particularly spending programs, where taxpayer dollars are most vulnerable to waste.

But the experts we consulted urged us to develop a set of companion tools that could be deployed to evaluate the effectiveness of existing programs. With budgets inevitably tight for the foreseeable future, they said, Washington urgently needs a better way to distinguish the most effective programs from those in need of reform, or risk cutting good programs simply because they have less political support.

We call these our "Reviewing What Works" tools. This section describes these questionnaires and how they can be incorporated into the federal government's emerging performance management system.

The problem

Congress and the executive branch have a long habit of layering new programs on top of existing ones. There are, for example, 110 funded programs on science, technology, and mathematics education across 14 departments. There are more than 100 programs that support youth mentoring across 13 agencies. [30]

As the president's fiscal commission recently reported: "The government will not be able to protect those in need or invest to achieve our nation's long-term potential growth if Washington squanders taxpayer dollars on duplicative programs with no measurable results."[31]

The Obama administration has made some progress in this area. It has already identified $20 billion in savings by reforming programs that were duplicative or

poorly performing. Congress has backed the administration on about 60 percent of the proposed savings, but these sums amount to just one-half of 1 percent of the federal discretionary budget.

Moreover, the processes for evaluating which existing programs work and which don't remain ad hoc. Congress requires all programs to collect performance information under the Government Performance and Results Act, but the information is often of poor (if voluminous) quality and there is scant evidence that agencies or budget officials use it to guide their decisions.

As part of the budget cycle, the administration has asked agencies to identify their least effective programs. But there is no common process across agencies to do this, and there are strong incentives for agencies to present as "least effective" those programs with the most political support—because leaders are unlikely to change them.

There are at least five reasons why Washington finds it so hard to evaluate existing programs.

First, even where programs are badly designed at the outset, they often become permanent because of pressure from the interest groups they benefit.

Second, program managers have no incentives to evaluate the effectiveness of their programs, only to defend them.

A third problem is that even where programs are being evaluated against initiative-specific goals, those goals may not be aligned with national priorities. Contributing to this ratcheting effect is the widely held view that any program that produces any benefit is worth keeping around, even if it's inefficient.

Fourth, there is a broader institutional issue: No one really "owns" the task of evaluating whether federal programs work.

Finally, the political process rewards people who come up with new ideas, not fix old ones. Interest groups court new policies, and reward politicians who champion their ideas. That means Washington decision makers tend to channel their energies into developing new policies rather than fixing existing programs.

"As president, I will go through the entire federal budget, page by page, line by line, and I will eliminate the programs that don't work and aren't needed. As for the programs we do need, I will make them work better and cost less."

– Sen. Barack Obama (D-IL), 2008

Recent developments in federal performance management

The Office of Management and Budget under George W. Bush developed the Program Assessment Rating Tool, a process to systematically evaluate all federal programs. Known as PART, the initiative showed initial promise but ultimately proved unsuccessful and was recently abandoned by the Obama administration.[32]

About a thousand government programs were subjected to PART scrutiny before it was discontinued. Under the process, OMB examiners rated programs for effectiveness based on agency answers to a list of 25 questions that focused on key aspects of good program design, such as whether there were clear outcome goals.

So why didn't this effort work? We think there are two key reasons. First, PART looked at each program on its own, which made it difficult to compare different programs across a policy area. The second key problem was that many agency officials felt that OMB examiners had too much power in deciding how a program fared under PART, that it was a process done *to* them rather than *with* them. Our Reviewing What Works tools and processes described below attempt to learn and build on the valuable PART experience.

There have been, since PART, important strides in federal performance management. The Obama administration has adopted 128 High Priority Performance Goals that define the key priorities for each agency, such as improving high-school graduation rates and doubling renewable energy generating capacity. These priority goals set out, for the first time, a list of major goals for government.

There have also been promising developments on the Hill. Congress recently passed the Government Performance and Results Modernization Act of 2010, which was signed by the president in January. The act requires the executive branch to work with Congress to adopt outcome-based government-wide goals and issue regular progress reports. These requirements address the key weaknesses of the 1993 Government Performance and Results Act, which has generated reams of program data, but little insight into outcomes. As a result, few programs have been using the information to enhance performance.

Solution: The Reviewing What Works process

Now that Congress and the administration have both required a set of goals for government, it becomes possible to evaluate programs against them. We believe that's how a system of evaluating existing program effectiveness should proceed. We think it will avoid the problems that the PART process encountered.

Consider for example the problem of homelessness. The administration has already adopted goals to end chronic homelessness and homelessness among veterans within five years, along with a commitment to end homelessness for youth, families, and children within 10 years.[33] It's now possible to evaluate the many programs that seek to reduce homelessness against their contribution to achieving these goals.

Here's how the Reviewing What Works tools would work across a policy area such as homelessness:

STEP 1: Form interagency panels for policy areas

The federal government should create and staff interagency panels around policy areas. Where similar panels already exist, such as the ones for homelessness and obesity, the Reviewing What Works process should be added to their purview. The same interagency panels that function as the neutral party in the Design for Success context could be used in this process.

STEP 2: Define common goals and compile inventories of programs

The interagency panels should define common goals across a policy area and compile an inventory of all government programs that contribute to these goals. That information gets included on a "policy strategy" form.

STEP 3: Complete "program effectiveness" questionnaire

Program managers answer a set of common questions on a "program effectiveness" questionnaire that helps identify whether the program works well.

STEP 4: Evaluate

Interagency panels review and assess "program effectiveness" responses on a separate "program evaluation" questionnaire.

STEP 5: Make informed decisions

Once all the programs across a policy area have been evaluated, the final versions of all three tools for each program—"policy strategy," "program effectiveness," and "program evaluation"—collectively inform decisions by executive and legislative branch officials about whether to continue, expand, reform, or end individual initiatives. (See Appendix 6 for the three tools)

This process is designed to address the shortcomings of the PART experience. In particular, we have been careful to ensure that interagency panels rather than OMB examiners assess the likely success of a new program.

Our approach looks at sets of programs across policy areas. We think that the effectiveness of existing programs requires an assessment of their contribution to a broad set of goals. This builds on the Obama administration's approach of setting High Priority Performance Goals for agencies.

Now that we've described how the process works, let's take a closer look at the kinds of questions the Reviewing What Works tools prompt officials to answer.

Reviewing What Works—the process

Step 1	Step 2	Step 3	Step 4	Step 5
Form an interagency panel for a policy area (such as homelessness)	Interagency panel defines goals across the policy area and lists programs using the Policy Strategy tool	Program managers fill out the Program Effectiveness tool for each program	Interagency panel considers all Program Effectiveness tools and fills out Program Evaluation tools for each program	Decision makers in executive and legislative branch use information to decide which programs merit expansion, reform, or elimination

We believe the questions for the Reviewing What Works tools should follow a similar format to the Design for Success tools. Our objective was to develop a set of questions that assessed the relative effectiveness of programs against the goals of government in a particular policy area. For each program, the Reviewing What Works tools ask questions in five areas:

Impact

We start by asking what impact the program is having on government goals in across the policy area. For example, what impact is a homelessness program having in reducing chronic homelessness?

Collaboration

How does the program interact with other initiatives? Is it duplicative of other efforts? Does it maximize the collective impact of similar programs?

Benchmarking

This section assesses the costs and outcomes of the program against "benchmark" indicators of similar initiatives. A good program is one that can demonstrate better outcomes or lower costs than other programs with similar goals.

Operational excellence

This section probes whether the program is well run. Have there been delays or cost overruns? Have stakeholders in state and local governments worked to support the program or obstruct it?

Adaptability

Finally, the questionnaire asks whether the program under evaluation has sought to learn from the experience—such as whether there are regular assessments of the program's progress toward achieving its goals.

These questions should lead respondents to paint a detailed picture of an existing program's effectiveness. To be sure, program managers may find some questions challenging to answer and try to avoid the most difficult ones.

That's why our proposed process works by evaluating programs across policy areas and against a common set of goals. And that's why the evaluation is led by inter-agency panels with expertise of the policy area drawn from across government.

Many participants will be program managers from other agencies. We believe that this process will create strong incentives for program managers to answer questions to the best of their abilities.

A time to act

As Washington budgets tighten, there's a risk that decisions about which programs to back and which ones to cut will be made with little understanding about which programs actually work. That's why there's an urgent need now to review the effectiveness of existing programs.

The Reviewing What Works tools and process provide a way forward. The processes are rooted in the new performance management framework emerging in both the Obama White House and in Congress.

But it won't be easy. Critics will complain that program managers should be focused on running programs, not answering questions about whether their programs work. But we have heard time and time again over the course of this project: We can no longer defer proper consideration of which programs are most and least effective. The time to act is now.

Conclusion

The central idea in this report is that there should be more prominent consideration during the design phase of whether a new government program is likely to work. We have also identified an urgent need for a formal system to evaluate which existing programs work and which do not.

To that end, we have designed a series of tools that would underpin new processes in the executive and legislative branch.

These are simple ideas, and the problems that we are trying to tackle are not new. Indeed, our Design for Success and Reviewing What Works tools ask fairly simple questions. And yet political considerations in both the legislative and executive branch context tend to squeeze out questions of effectiveness and impact.

And so these questions don't get the prominence they deserve, either before or after programs are launched.

So what needs to happen to turn the ideas in this report into a reality? One of the hallmarks of a successful initiative is that it should be tried out before it is rolled out at any significant scale. For that reason, we believe that the processes and tools set out in this report should be tested on a small scale before any widespread adoption.

In the executive branch, the processes for new and existing program might be piloted across three or four policy areas. In Congress, a handful of committees might implement a version of the tools and monitor the impact they have on the decision-making process.

States would also be fertile testing grounds. More than half of U.S. states will swear in a new governor in 2011, and most governors are working with historically tight budgets in the coming years. The tools in this report would help them evaluate the relative effectiveness of existing programs, and determine which new program ideas are most likely to succeed.

Appendix 1

Design for Success: A checklist

Design for Success Program Checklist

A. Are the goals clear and the cost estimates accurate?

1. Clear understanding of the problem in measurable terms ❏
2. Clear goals expressed as outcomes over a defined time scale ❏
3. Accurate cost estimates ❏
4. Sufficient funds to achieve goals ❏

B. Is this the right approach?

5. Fills a clear gap in current government services ❏
6. Cannot achieve goals through expansion or modification of existing programs ❏
7. Little duplication or overlap with current federal programs ❏
8. Reasonable conclusion that this approach is the best option ❏
9. Reasonable total cost for expected outcomes compared to alternatives ❏

C. Why will the program work?

10. Reasonable expectation that key stakeholders will alter their ways of working or behavior ❏
11. Professional, independent research indicates approach will work ❏
12. Program expands on what worked on the smaller scale, or plan to test before rolling out on the large scale ❏

D. Does the program establish the right incentives?

13. Establishes clear incentives that are aligned with program goals ❏
14. Beneficiaries reasonably expected to change behavior ❏
15. Beneficiaries have been consulted and indicate that the program is likely to work ❏
16. Strong understanding of areas most prone to gaming, and risks mitigated in program design ❏

E. How will the program be implemented?

17. Reasonable timeline for program rollout ❏
18. Implementers have bought into the program and are prepared to carry it out ❏
19. Sufficient plan for hiring, training, and IT development ❏
20. Reasonable procurement strategy and sufficient staff to negotiate and manage contracts ❏
21. Design takes into account views from key outside stakeholders ❏
22. Clear strategy for minimizing impact on negatively affected parties ❏
23. Reasonable effort to minimize unintended consequences ❏
24. Strong risk-mitigation strategy in place ❏

F. How will you monitor success and rethink the approach?

25. Key indicators of success identified for each goal ❏
26. Clear plan for collecting timely, accurate data that reflects outcomes ❏
27. Takes advantage of existing data systems where possible ❏
28. Coordinates data collection with other programs where possible ❏
29. Reasonable plan for rethinking approach ❏
30. Low sunk costs ❏
31. Can be quickly phased out or terminated if deemed ineffective or no longer necessary ❏

Appendix 2

Design for Success tools for the executive branch

Design for Success **Executive** Tool—Program Details

Description

This tool is intended for new spending programs that are created in the executive branch and normally funded through the annual budget and appropriations process. For this tool, proponents of a program respond to each question in clear, concise narrative (approximately 200 words per response), with the opportunity to attach additional details as appendices. An interagency panel on the specific policy area considers the information in this form to assess the program's likelihood of success in the Program Assessment Tool.

Program name: _____ **Reference #:** _____

Agency: _____ **Bureau:** _____

Author: _____ **Approved by:** _____ **Date:** _____

A. Program description

1. Describe the proposed program and the problem it is trying to address. Also lay out the current condition of the problem in measurable terms.

2. What do you expect the program to achieve? Lay out the program's specific outcome goals in measurable terms over a defined time scale. Be clear about what the program will do for the American people, not simply its expected outputs and expenditures.

3. What will be the program's startup costs to the federal government? $ _____

What will be the recurring annual costs to the federal government? $ _____

4. How do you know that the cost estimates are as accurate as possible for the structure and lifespan of the program? Explain how you came up with this estimate, and why you think the amount is enough to carry out the program's objectives. If possible, provide the error range or confidence interval for these cost estimates.

B. Policy approach

5. Is this the right approach?

5A. Explain why there needs to be a new program to address this problem. Explain why the private sector, states, or local governments can't address the problem without a federal program. Why can't it be addressed by expanding or modifying an existing program (either within the agency, in other agencies, or other levels of government), or through a regulatory or tax approach?

Design for Success **Executive** Tool—Program Details

5B. What type of program is this (direct federal program, competitive or formula/block grant, voucher program, credit program, etc.)? Why did you choose this type of program? Discuss the alternative approaches that might achieve the same outcome, and why they were rejected. Why is this particular policy approach the best way to address the problem?

5C. Describe any current federal programs (including those in other agencies) that are similar in design to the program or intend to achieve similar outcomes. How is your program different from existing ones? How will your program leverage, coordinate with, or replace these programs to prevent overlap and duplication?

5D. Explain why this is an efficient use of federal government funds. How do the cost estimates compare to alternative approaches, considering the likely impact? Have estimates of costs and impact associated with the program been independently validated?

6. Why will the program work?

6A. How will the program achieve its outcomes? Describe which stakeholders need to alter their ways of working or behavior in order for the program to succeed, and why you believe they will do so.

6B. What evidence suggests that this program is likely to achieve its outcome (academic research, GAO and other government reports, independent surveys, etc.)? How do you know this evidence is reliable?

6C. Has this program been implemented elsewhere (pilot programs, states, localities, other countries, etc.)? If so, describe the locations, effectiveness, and lessons learned. Have randomized control trials or similarly rigorous evaluations of impact been conducted?

6D. If the program has not been implemented elsewhere, discuss how you plan to test the new idea. If this program has been implemented elsewhere, discuss how you know that you are expanding the model based on what has worked, and that it will succeed at the scale you propose.

7. Does the program establish the right incentives?

7A. Describe the incentive structure put in place to achieve the program's goals. If the program tries to impact or change behavior, how will the targeted individuals be compelled to make that change? Have the expected beneficiaries of the program been consulted to determine that it will work from their perspective?

7B. Discuss any foreseeable ways to cheat the system. How does the program design address these risks, and how do you plan to minimize their occurrence?

C. Implementability

8. How will the program be implemented?

8A. Briefly describe the expected timeline for implementing the program. Why do you think this timeline is realistic?

Design for Success **Executive** Tool—Program Details

8B. Discuss the extent to which the agency is ready to administer the program. How do you plan to secure the necessary staff (including management and program staff), skill base, and IT infrastructures?

8C. Who will be responsible for implementing the program at the state and/or local level? Have they been consulted, and are they committed to and prepared for the program?

8D. Identify groups that can positively and negatively affect the program's success (other federal agencies, unions, interest groups, business groups, other stakeholders). Have these groups been consulted, and have their views been taken into account in the program design?

8E. Discuss the program's expected procurement needs. Are the expected procurement costs realistic? Are the necessary contract staff and review structures in place to ensure that procurement costs are minimized and terms are closely monitored?

8F. Describe the major risks for the program (such as possible occurrences that can impede success of the program) and where they come from. For each major risk, discuss the impact if it should occur, the probability of its occurrence, and mitigation strategies.

8G. Do you foresee any negative effects of the program (for example, on people or firms)? If so, what are they? What are the possible unintended negative consequences of the program? How do you plan to minimize the adverse impact of these consequences?

9. How will you monitor success and rethink the approach?

9A. List the key performance indicators through which you intend to track the program's outcomes.

9B. Is the data you plan to collect an accurate and timely reflection of the program's performance toward achieving the outcome? To the extent possible, are you coordinating data collection with other programs?

9C. Lay out the plan for evaluating the program's progress toward achieving its goals. How often will performance assessments take place and to whom will progress be reported? How will you assess whether the program, and not other factors, is driving any observed changes in the problem? Will you be able to tweak your approach along the way as you learn what is and is not working?

9D. Describe the plan put in place for rethinking, phasing out, or terminating the program if its progress is off course or no longer deemed necessary.

Design for Success **Executive** Tool—Program Assessment

Description

For this tool, an interagency panel assesses the proposed program's likelihood of success if implemented. The panel probes the agency on the contents of the Program Details tool and validates the information on it. The panel then assesses each key success factor on a scale of 1-10 (where 1 is very weak and 10 is very strong) and provides a brief explanation of each assessment. The agency in charge of implementing the proposed program then submits the completed Program Assessment tool and the validated Program Details tool to OMB as part of its annual budget request. This information accompanies the program proposal throughout the budget and appropriations process.

Program: _____ **Reference #:** _____

Agency: _____ **Bureau:** _____

Policy area: _____ **Interagency panel:** _____ **Date:** _____

A. Program description

1. Describe the proposed program and the problem it is trying to address. Also lay out the current condition of the problem in measurable terms.

2. What do you expect the program to achieve? Lay out the program's specific outcome goals in measurable terms over a defined time scale. Be clear about what the program will do for the American people, not simply its expected outputs and expenditures.

3. What will be the program's startup costs? $ _____

What will be the program's recurring annual costs? $ _____

4. How do you know that the cost estimates are as accurate as possible for the structure and lifespan of the program? Explain how you came up with this estimate, and why you think the amount is enough to carry out the program's objectives. If possible, provide the error range or confidence interval for these cost estimates.

B. Policy approach

5. Is this the right approach?

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Very weak   Weak   Neither weak nor strong   Strong   Very strong
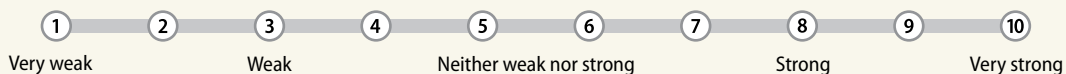
Weak approach

No clear need for a new program; high duplication/overlap with current federal programs; unclear reasons behind specific policy approach; little consideration of alternative approaches; excessive costs for the expected outcomes compared to alternatives

Strong approach

Fills a clear gap in current government services; cannot achieve goals through expansion or modification of existing programs; little duplication or overlap current federal programs; reasonable conclusion that this approach is the best option; reasonable total cost for expected outcomes compared to alternatives

Reasons for assessment

6. Why will the program work?
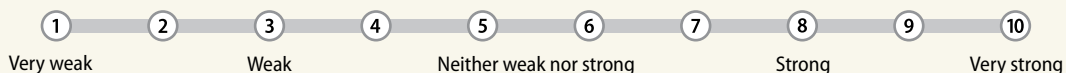
(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak evidence

Unreasonable expectation that key stakeholders will alter their ways of working or behavior; similar programs never implemented elsewhere; research on the approach either unavailable or concludes it is ineffective; no plan to test the program before rolling out on the large scale/expanding on the model too quickly

Strong evidence

Reasonable expectation that key stakeholders will alter their ways of working or behavior; professional, independent research indicates approach will work; program expands on what worked on the smaller scale/plan to test before rolling out on the large scale

Design for Success **Executive** Tool—Program Assessment

Reasons for assessment

7. Does the program establish the right incentives?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak incentive system	*Strong incentive system*
Incentives are unclear or not aligned with program goals; program will not reasonably compel beneficiaries to change behavior; beneficiaries not consulted to determine that the program is likely to work; foreseeable ways to cheat the system not addressed in program design	Establishes clear incentives that are aligned with program goals; beneficiaries reasonably expected to change behavior; beneficiaries have been consulted and indicate that the program is likely to work; strong understanding of areas most prone to gaming and risks mitigated in program design

Reasons for assessment

C. Implementability

8. How will the program be implemented?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak implementation plan	*Strong implementation plan*
Unreasonable timeline for program rollout; implementers not bought in or views not sought; insufficient plan for hiring, training, and IT development; unrealistic procurement strategy and insufficient staff to negotiate and manage contracts; views of key outside stakeholders not taken into account; negatively affected parties and unintended consequences not acknowledged and no plan for minimizing impact; little consideration of program risk	Reasonable timeline for program rollout; implementers have bought into the program and are prepared to carry it out; sufficient plan for hiring, training, and IT development; reasonable procurement strategy and sufficient staff to negotiate and manage contracts; design takes into account views from key outside stakeholders; clear strategy for minimizing impact on negatively affected parties; reasonable effort to minimize unintended consequences; strong risk mitigation strategy in place

Reasons for assessment

9. How will you monitor success and rethink the approach?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak   Weak   Neither weak nor strong   Strong   Very strong
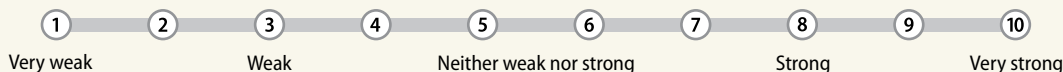
Weak assessment/contingency plan	*Strong assessment/contingency plan*
No indicators for success; no plan to collect timely, accurate data that reflects outcomes; no plan to coordinate data collection with other programs; unclear or insufficient reporting schedule; difficult to rethink or tweak the program after implementation; high sunk costs; difficult to terminate or requires long phase-out process if unsuccessful	Key indicators of success identified for each goal; clear plan for collecting timely, accurate data that reflects outcomes; takes advantage of existing data systems where possible; coordinates data collection with other programs where possible; reasonable plan for rethinking approach; low sunk costs; can be quickly phased out or terminated if deemed ineffective or no longer necessary

Reasons for assessment

The current process for adopting new spending programs in the executive branch

New program ideas in the executive branch are largely developed within individual agencies.[34] If the agency thinks a new idea is good enough for budgeting consideration they include the program in their annual budget request to the Office of Management and Budget, or OMB. Through a series of reviews, passbacks, and revisions, OMB determines whether to include the new idea in the president's budget. But the final funding decision is made in Congress through appropriations legislation.[35]

Each new program idea receives considerable scrutiny throughout this process. Agency budget shops assess the costs and merits of all new program ideas before requesting funds. Budget examiners at OMB are charged with reviewing each activity in an agency's budget request, and they identify which programs are worthy of federal funds. The problem is that often serious design flaws fall through the cracks and critical questions simply go unasked.

Proponents of new programs present their ideas in the most flattering possible light, while budget examiners and other key decision makers do their best to address essential aspects of the program's design. They may well probe whether the policy is the right one, but we heard that considerations of the implementability of new ideas were almost entirely absent.

There is currently no formal or common set of questions that agencies ask internally or are asked by OMB.[36] The Design for Success review is one way to institute such a process.

The Design for Success process for the executive branch

Step #1: The proponent uses the "program checklist" to design an initiative

When an agency comes up with an idea for a new government initiative, the program proponent—likely an agency staffer—would use the "program checklist" tool to expose potential gaps in the proposal that need to be addressed. The proponent uses the checklist as a benchmark for good program design and tweaks the program approach along the way.

Step #2: The proponent completes the "program details" questionnaire

Once the program has been designed and is ready for further consideration the proponent completes the details tool and submits the idea through the agency's internal review process. Agency leaders, advised by their budget and performance shops, would use information on the details tool to help them determine which new program ideas are most likely to succeed.

Step #3: A neutral party completes the "program assessment" questionnaire

Agency staff and other government officials told us a neutral party would have to validate any information presented on the program details tool. We heard that a neutral party should summarize the information submitted by the program manager into a simpler, more accessible form—and in doing so provide an assessment of whether the program is likely to succeed. This is what the program assessment tool is for.

We debated which government body should undertake the role of the neutral party. We considered asking each agency to convene a neutral panel to assess new proposals and to complete program assessment forms for all new proposals in their agencies. But we rejected that approach for two reasons. First, as we enter a period of fiscal tightness, we expect the number of new programs to go down considerably. There will be some new programs, but there may not be enough to justify each agency setting up its own separate process. Second, one weakness of the current system is that duplication often emerges between programs at different agencies because they are developed in silos.

Another option was to ask OMB examiners to lead the assessment process. In reality, though, they have little direct expertise in managing programs. The examiners are not always well placed to know whether a program has the necessary ingredients for success. And as the PART experience demonstrated other agencies do not see OMB as neutral. So it made more sense to have an interagency process that includes some external participants.

We concluded that the best way to independently assess new ideas is through review panels focused on particular policy goals such as reducing homelessness, promoting public health, improving school attainment, or improving competitiveness. The review panels could be convened by a lead agency, and they would include budget and performance experts from the key agencies, the relevant OMB examiners, and program managers with direct experience of administering gov-

ernment programs. The panels also would benefit from some outside involvement such as state and local government employees who are responsible for implementing government programs, notable academics, and other experts on the issue.

Together this group would be charged with reviewing the information program proponents present and assessing the likelihood of a new initiative's success based on that information. They would summarize the information on the program assessment tool, which is designed to be short and easy to access for busy decision makers. It would be submitted by agencies to OMB along with budget requests for any new programs.

Step #4: Decision makers use the information on both questionnaires to guide their scrutiny

Once both forms have been completed, agencies will make final decisions on whether they wish to include the program idea in their budget submission to OMB. They may want to tweak the proposal in the light of the assessment by the interagency panel.

For proposals that are included in agency submissions, OMB staff should consider the information in both forms when deciding whether a program should be included in the president's budget. We envision that these forms will be a critical resource during the pass-back and budget drafting process—as OMB asks program designers to make changes to maximize the chances of an idea's success.

Decision makers will want to review the tools' information alongside other considerations, such as the magnitude of the issue that the program addresses, the urgency with which that issue needs to be tackled, the constituency of support for the new idea, and how the program will play in the political arena. The tools do not cover this ground intentionally as there are already strong pressures in the system to address these issues.

If OMB and the White House decide to include the program in the president's budget, the information from both tools would be transmitted to Congress by the administration along with the final budget documents. This information will help appropriators consider the program's likelihood of success. [37]

The issue of transparency was the topic of much debate while developing our tools. We believe that the information on these tools should be available publicly for all new programs in the president's budget. If a proposal does not make it into the budget the information should be treated as pre-decisional and not be published.

There is one potential downside to this level of transparency, however. Some government officials expressed concern that making the assessment tool information public could affect the tool's legitimacy. The administration might feel compelled to present all new program ideas in too rosy a light, while they might be more honest about design flaws in an internal government document. This is a serious concern—honesty in the details tool is critical to the Design for Success process. But we have tried to frame questions in such a way that readers would be able to tell if the answers amount to fabrications or exaggerations. And the interagency review should also act as a check in this regard.

Appendix 3

Design for Success tools for the legislative branch

Design for Success **Legislative** Tool—Program Details

Description

This tool is intended for new spending programs that are created through legislation. For this tool, the program proponent responds to each question in clear, concise narrative (approximately 150 words per response), with the opportunity to attach additional details as appendices. Where it proves difficult for a bill sponsor to answer a question (for example, when the details of the program are to be resolved by the agency responsible for administering the program), they should say so. A legislative review body (either the committee under which appropriate jurisdiction falls or an independent organization like the Congressional Budget Office) considers the information in this form to assess the program's likelihood of success in the Program Assessment Tool.

Program name: _____ Bill #: _____

Sponsor: _____ Committee: _____ Date: _____

A. Program description

1. Describe the proposed program and the problem it is trying to address. Also lay out the current condition of the problem in measurable terms.

2. What do you expect the program to achieve? Lay out the program's specific outcome goals in measurable terms over a defined time scale. Be clear about what the program will do for the American people, not simply its expected outputs and expenditures.

3. In advance of the Congressional Budget Office score, what do you expect the program to cost the federal government annually? (Provide a range if necessary.) $ _____

4. Explain how you came up with the cost estimate, and why you think it is accurate.

B. Policy approach

5. Is this the right approach?

5A. Explain why there needs to be a new program to address this problem. Explain why the private sector, states, or local governments can't address the problem without a federal program. Why can't it be addressed by expanding or modifying an existing program (either federal, state, or local), or through a regulatory or tax approach?

Design for Success **Legislative** Tool—Program Details

5B. What type of program is this (direct federal program, competitive or formula/block grant, voucher program, credit program, etc.)? Discuss the alternative approaches that might achieve the same outcome, and why they were rejected.

5C. Describe any current federal programs that are similar in design to the program or intend to achieve similar outcomes. How are you avoiding overlap and duplication?

6. Why will the program work?

6A. How will the program achieve its outcomes? Describe which stakeholders need to alter their ways of working or behavior in order for the program to succeed, and why you believe they will do so.

6B. What evidence suggests this program is likely to achieve its outcome (academic research, GAO and other government reports, independent surveys, etc.)? How do you know this evidence is reliable?

6C. Has this program been implemented elsewhere (pilot programs, states, localities, other countries, etc.)? Discuss how you know that you are expanding the model based on what has worked, and that it will succeed at the scale you propose. If the program has not been implemented elsewhere, discuss how you plan to test the new idea.

7. Does the program establish the right incentives?

7A. Describe the incentive structure put in place to achieve the program's goals. If the program tries to impact or change behavior, how will the targeted individuals be compelled to make that change? Have the expected beneficiaries of the program been consulted to determine that it will work from their perspective?

Design for Success **Legislative** Tool—Program Details

7B. Discuss any foreseeable ways to cheat the system. How does the program design address these risks, and how do you plan to minimize their occurrence?

C. Implementability

8. How will the program be implemented?

8A. Briefly describe the expected timeline for implementing the program. Why do you think this timeline is realistic?

8B. Which agency or bureau will be responsible for administering the program? Who will be responsible for implementing the program at the state and/or local level? Have these groups been consulted, and are they committed to and prepared for the program?

8C. Describe the major risks for the program (such as possible occurrences that can impede success of the program). Also identify groups that can positively and negatively affect the program's success (federal agencies, unions, interest groups, business groups, and other stakeholders). Have these groups been consulted, and have their views been taken into account in the program design?

8D. Do you foresee any negative effects of the program (for example, on people or firms)? If so, what are they? What are the possible unintended negative consequences of the program?

8E. What are the potential obstacles to phasing out or terminating the program if it proves ineffective at achieving the desired outcome or is no longer deemed necessary?

Design for Success **Legislative** Tool—Program Assessment

Description

For this tool, a legislative review body—either the committee under which appropriate jurisdiction falls or an independent organization like the Congressional Budget Office—assesses the proposed program's likelihood of success if implemented. The reviewers probe the bill's sponsor on the contents of the Program Details tool and validate the information on it. The reviewers then assess each key success factor on a scale of 1-10 (where 1 is very weak and 10 is very strong), and provide a brief explanation of each assessment. Members of the committee then consider the information in the Program Assessment tool and the validated Program Details tool during the committee hearing. If the committee approves the bill for a floor vote, both tools accompany the bill for the remainder of its lifespan.

Program: _____ **Bill #:** _____

Policy area: _____ **Interagency panel:** _____ **Date:** _____

A. Program description

1. Describe the proposed program and the problem it is trying to address. Also lay out the current condition of the problem in measurable terms.

2. What do you expect the program to achieve? Lay out the program's specific outcome goals in measurable terms over a defined time scale. Be clear about what the program will do for the American people, not simply its expected outputs and expenditures.

3. In advance of the Congressional Budget Office score, what do you expect the program to cost the federal government annually? (Provide a range if necessary.) $ _____

4. Explain how you came up with this estimate, and why you think it is accurate.

B. Policy approach

5. Is this the right approach?

①————②————③————④————⑤————⑥————⑦————⑧————⑨————⑩

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak approach

No clear need for a new program; high duplication/overlap with current federal programs; unclear reasons behind specific policy approach; little consideration of alternative approaches

Strong approach

Fills a clear gap in current government services; cannot achieve goals through expansion or modification of existing programs; little duplication or overlap with current federal programs; reasonable conclusion that this approach is the best option

Reasons for assessment

Design for Success **Legislative** Tool—Program Assessment

6. Why will the program work?
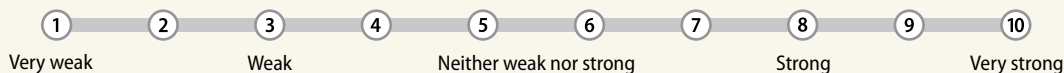
1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak · Weak · Neither weak nor strong · Strong · Very strong

Weak evidence
Unreasonable expectation that key stakeholders will alter their ways of working or behavior; similar programs never implemented elsewhere; research on the approach either unavailable or concludes it is ineffective; no plan to test the program before rolling out on the large scale/expanding on the model too quickly

Strong evidence
Reasonable expectation that key stakeholders will alter their ways of working or behavior; professional, independent research indicates approach will work; program expands on what worked on the smaller scale/plan to test before rolling out on the large scale

Reasons for assessment
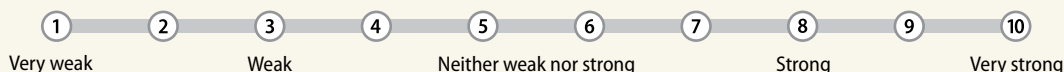
7. Does the program establish the right incentives?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak · Weak · Neither weak nor strong · Strong · Very strong

Weak incentive system
Incentives are unclear or not aligned with program goals; program will not reasonably compel beneficiaries to change behavior; beneficiaries not consulted to determine that the program is likely to work; foreseeable ways to cheat the system not addressed in program design

Strong incentive system
Establishes clear incentives that are aligned with program goals; beneficiaries reasonably expected to change behavior; beneficiaries have been consulted and indicate that the program is likely to work; strong understanding of areas most prone to gaming and risks mitigated in program design

Reasons for assessment

C. Implementability

8. How will the program be implemented?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak · Weak · Neither weak nor strong · Strong · Very strong
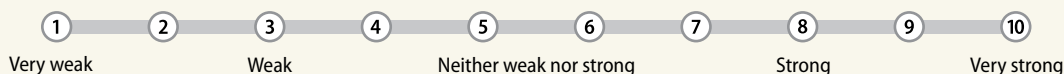
Weak implementation plan
Unreasonable timeline for program rollout; administrating agency/bureau and state/local implementers not bought in or views not sought; views of key outside stakeholders not taken into account; negatively affected parties and unintended consequences not acknowledged; little consideration of program risk; high sunk costs; difficult to terminate or requires long phase-out process if unsuccessful

Strong implementation plan
Reasonable timeline for program rollout; administrating agency/bureau and state/local have bought into the program and are prepared to carry it out; design takes into account views from key outside stakeholders; clear understanding of negatively affected parties; reasonable effort to minimize unintended consequences; strong risk-mitigation strategy in place; low sunk costs; can be quickly phased out or terminated if deemed ineffective or no longer necessary

Reasons for assessment

The current process for adopting new legislative programs

Federal programs are often created directly through legislation. A congressional member thinks of a new government initiative, drafts a piece of legislation that establishes (and sometimes directly funds) a new government program, and submits the bill to the appropriate committee or subcommittee for a hearing. If the committee decides that the bill is ready it's passed on to a floor vote. After passing, it's sent to a committee in the opposite chamber, approved for another floor vote, and eventually sent to the president's desk. Congress has effectively created a new government program when the bill is signed into law.[38]

The exact process for reviewing a new piece of legislation varies from committee to committee. But Hill staffers explained to us that no one had a set of questions they used when developing a new program or that they asked of every new government initiative. That means that no formal mechanism exists in Congress to ensure that essential questions of duplication, evidence, and implementation are asked during the debate.

The Congressional Budget Office's scoring process for bills is basically a review process in the legislative branch. Every bill brought to the floor in either chamber of Congress must receive a CBO "score," or a technical assessment of a new bill's likely cost. This helps ensure that proponents do not underestimate costs in order to win support for their idea.

The CBO score is a crucial part of the legislative process. But it does not go far enough. The score only looks at budgetary impact, which ignores whether the program is likely to work. Our proposed tools address that weakness. Congress needs a rigorous process that holds all new program ideas to the appropriate level of scrutiny before members are expected to vote on it.

The Design for Success process for the legislative branch

Step #1: The proponent uses the "program checklist" to design an initiative

When a congressional member decides he would like to propose a new program, he will likely ask one of his staff to lead on its development. The staffer uses the checklist as a benchmark for good program design while developing the program. The "program checklist" tool helps them get the proposal right.

Step #2: The proponent completes the "program details" questionnaire

When the bill is drafted the lead sponsor of that bill should be responsible for formally completing the details tool. [39]

The information from the program details tool should go with the bill when it is submitted to committee.[40] This provides committee staff with essential information on the design of any program created by the new legislation.

Step #3: Complete Program Assessment Tool

The program assessment tool would be completed during the committee review phase. Ideally, the assessment will be completed prior to the committee hearing, providing decision makers with as much information as possible before they deliberate on the bill so that they can seek changes to improve the program design.

Similar to the interagency assessment we propose for programs in the executive branch, it is important that an independent entity conducts the assessment of legislative programs. One option is that the committee staff could assess a proposal's likelihood of success as part of its regular review process. Another option would require an independent organization with substantive expertise, such as the Congressional Budget Office, to consider key components of the program's design, using information provided by bill's sponsors.

There are pros and cons to both options. Many see questions of likely success as too subjective for an organization such as CBO. Equally, there is a risk that committee members find it hard to put aside partisan views when considering program proposals. Selecting the body best fit to carry out this assessment should be left to committees.

Step #4: Decision makers use the information on both questionnaires to guide their scrutiny

The information on both questionnaires should follow the bill for the remainder of its lifespan. If the bill makes it to a floor debate the information in the tools will help guide the discussion and allow legislators to consider its likelihood of success.

Decision makers will want to review the tools' information alongside other considerations, such as the magnitude of the issue that the program addresses, the urgency with which that issue needs to be tackled, the constituency of support for the new idea, and how the program will play in the political arena. The tools do not cover this ground intentionally as there are already strong pressures in the system to address these issues.

As with the executive tools, it's difficult to determine the appropriate level of transparency for these tools. Some congressional staffers expressed concern that publishing this information could deter honesty about flaws in the program—and that bill sponsors would always feel the need to claim their program is perfect. We attempted to frame the questions so that readers would be able to tell if the answers were exaggerations. And the independent review process should also provide a check against such behavior.

Appendix 4

A draft tool for proposed regulations

Design for Success Tools—Regulatory proposals (draft)

Description

This draft tool is intended for regulatory proposals (including those that amend existing regulations). For this tool, proponents of a regulation (i.e. agency staff) respond to each question in clear, concise narrative (approximately 200 words per response), with additional details attached as appendices. This form summarizes some of the information that would be form part of the Regulatory Impact Analysis, and would be considered by the Office of Information and Regulatory Affairs in OMB alongside a completed RIA.

Program: _____   **Date:** _____

Agency: _____   **Bureau:** _____

A. Regulatory proposal—costs and benefits

1. Briefly explain the proposed regulation in clear and nontechnical language.

2. Describe the problem it is trying to address. Also lay out the current condition of the problem in measurable terms.

3. What do you expect the regulation to achieve? Lay out the outcomes that you expect to result from the regulation in measurable terms over a defined time scale. Be clear about what the regulation will do for the American people (for example, a reduction in annual vehicle fatalities by 1 percent), not simply the number of people or firms affected.

4. Lay out (in tabular form) the monetized costs and benefits of the proposed regulation following OMB guidance for Regulatory Impact Analysis. Also include any nonmonetized costs or benefits.

5. Are estimates of costs and benefits based on the best reasonably obtainable scientific, technical, and economic information? Explain how you came up with the estimates and whether they have been independently verified.

6. What costs or benefits were you unable to quantify and monetize? Explain why you were unable to do this.

7. What are the likely costs to federal, state, or local governments of the new regulation? Where do these costs fall and over what time period?

Design for Success Tools—Regulatory proposals (draft)

B. Policy approach

8. Why is this the right approach?

8A. Explain why there needs to be a regulatory change to address this problem. Explain what would happen in the absence of a regulatory change.

8B. Could another policy approach such as a tax measure or spending program achieve the goals of the regulation? If so, were they considered and why were they rejected?

8C. Describe other federal regulations that are similar in design or intend to achieve similar outcomes. To the extent possible, are you leveraging existing regulations rather than creating new regulatory approaches?

8D. Discuss the alternative regulatory approaches that you considered. Explain why they were rejected, including the costs and benefits of different approaches.

8E. Explain why you believe that the benefits justify the costs.

9. Why will the program work?

9A. What evidence (e.g., academic research, GAO and other government reports, independent surveys, etc) suggests that this regulatory approach is likely to achieve its outcomes? How do you know that this evidence is reliable?

9B. Has a similar regulatory approach been implemented elsewhere (states, localities, other countries, etc)? If so, describe the locations, effectiveness, and lessons learned. Have rigorous evaluations of impact been conducted?

Design for Success Tools—Regulatory proposals (draft)

9C. Have those affected by the program been consulted? What are their views on its likely success? Have they made suggestions on how to improve the program? How have you adjusted the program design to take account of these comments?

10. Does the program establish the right incentives?

10A. How will the regulatory program achieve its outcomes? Describe who needs to alter their behavior in order for it to succeed, and why you believe they will do so.

10B. What level of compliance do you expect when the regulation comes into force?

10C. Discuss any foreseeable ways to cheat the system. How does the program design address these risks, and how do you plan to minimize their occurrence?

C. Implementability

11. How will the program be implemented?

11A. Briefly describe the expected timeline for implementing the regulation. Why do you think this timeline is realistic?

11B. Will a federal agency be responsible for enforcement of the regulation? Discuss the extent to which it is ready to do so. How does it plan to secure the necessary staff, skill base, and IT infrastructures?

Design for Success Tools—Regulatory proposals (draft)

11C. Will the program be enforced by states or localities? Have they have been consulted, and are they committed to the program? In particular, do they have the necessary budget, staff, and IT infrastructures?

11D. Describe the major risks for the regulatory program (such as possible occurrences that can impede success of the program) and where they come from. For each major risk, discuss the impact if it should occur, the probability of its occurrence, and mitigation strategies.

11E. Do you foresee any negative effects of the regulatory program (for example, on people or firms)? If so, what are they? What are the possible unintended negative consequences of the program? How do you plan to minimize the adverse impact of these consequences?

12. How will you monitor success and rethink the approach?

12A. List the key performance indicators through which you intend to track the regulatory program's outcomes.

12B. Is the data you plan to collect an accurate and timely reflection of the regulation's performance toward achieving the outcome? To the extent possible, are you coordinating data collection with other programs (including nonregulatory programs)?

12C. Lay out any plan for evaluating the regulation's progress toward achieving its goals. Will you be able to tweak your approach along the way as you learn what is and is not working?

12D. What are the possible obstacles to phasing out or terminating the regulation if it proves ineffective at achieving the desired outcome or is no longer deemed necessary?

Appendix 5

A draft tool for proposed tax provisions[41]

Design for Success Tools—Tax Provisions (draft)

Description

This draft tool is intended for proposed tax expenditures. It would be completed for tax provisions (including those that modify existing provisions) that provide incentives or disincentives for certain activities (such as saving for retirement). It would not be completed where the objective of the provision is entirely tax related (for example, where a change to the tax law aims to increase the accuracy of income measurement or make broad changes in tax rates). For this tool, the proponent of the provision responds to each question in clear, concise narrative (approximately 150 words per response) with the opportunity to attach additional details as appendices.

Name of proposed tax provision: _____ **Bill #:** _____

Sponsor: _____ **Committee(s):** _____ **Date:** _____

A. Description—costs and benefits

1. Describe the proposed tax provision.

2. In what form is the tax incentive or subsidy delivered? Please circle:

Exemption—a provision that exempts or excludes a certain kind of income from taxation

Deduction—a provision that permits taxpayers to deduct a certain expense from the amount of income that is subject to tax

Rate preference—a provision that provides lower rates for income earned in certain ways

Credit—a provision that directly reduces the taxpayer's tax liability after the determination of income and the application of tax rates

Deferral feature—a provision that shifts tax liabilities to later years

3. What activity does the tax provision seek to discourage or encourage (for example, saving for retirement, getting homes insulated, consuming tobacco products)?

4. What is the current level of the activity? Is it too high or not high enough?

5. What impact do you believe the tax provision will have on that level and over what timeframe?

6. What are the benefits to the public that would result from the change?

7. In advance of the official revenue estimate by the Joint Committee on Taxation, what do you expect the revenue effects of the tax provision to be? (Provide a range if necessary.)

$ _____ over _____ years

Design for Success Tools—Tax Provisions (draft)

8. Explain how you came up with the estimate and why you think it is reasonable.

9. If the new program provides a tax benefit, is the fiscal cost commensurate with the policy goal? If the new program imposes a new form of tax, is the cost to taxpayers commensurate with the policy goal?

10. Are the revenue effects likely to vary significantly between years over a 5-, 10-, or 20-year time horizon? If so, please describe the nature of the likely variation.

B. Policy approach

11. Why is this the right approach? Why choose a tax measure?

11A. Explain why there needs to be a new initiative at the federal level to encourage or discourage the activity. Explain why the private sector, states, or local governments can't take action instead.

11B. Rather than adopting a tax provision, why wouldn't it be better to expand or modify an existing spending program or regulatory measure?

11C. Why do you think that a new tax provision is a better way forward than a new spending program? In particular, are there administrative advantages to running the program through tax returns?

Design for Success Tools—Tax Provisions (draft)

11D. Do all (or almost all) of the intended beneficiaries file tax returns? If not, why are you not proposing a spending measure?

11E. Tax information is generally confidential while the beneficiaries of spending programs (contracts, grants) are generally made public. Explain whether important transparency is lost by making the program a tax expenditure.

11F. Is the information required to determine eligibility already collected by the IRS? If not, explain why it would be relatively easy to collect it.

11G. Describe any current federal spending programs or tax provisions that intend to achieve similar outcomes. How are you avoiding overlap and duplication?

11H. Is it desirable to sunset the provision or does it need to be permanent to meet its objectives? Explain the reasons for your answer.

12. Why will the provision work?

12A. Is there any evidence (academic research, GAO and other government reports, independent surveys, etc.) that suggests the provision is likely to achieve the change in activity level? How do you know this evidence is reliable?

12B. Has a similar tax provision been implemented elsewhere (states, localities, other countries, etc.)? Discuss how you know you are expanding the model based on what has worked, and whether it will succeed at the scale you propose.

Design for Success Tools—Tax Provisions (draft)

13. Does the provision establish the right incentives?

13A. If it is a deduction, exemption, or deferral feature, are there policy reasons to provide a different subsidy in dollar terms to taxpayers in different tax brackets?

13B. If it is a deduction for individuals, is it an itemized or nonitemized deduction? What are the administration and/or policy reasons it should be itemized or nonitemized?

13C. If it is a credit, explain whether or not it is refundable. If it is not refundable, explain why the incentives behind the policy do not apply to taxpayers without federal income tax liability.

13D. Given the structure of the tax provision, is it targeted to those taxpayers who will be most responsive to the incentives or to those who are most in need of its benefits?

13E. Do you foresee taxpayers (e.g. individuals or firms) who would lose out as a result of the tax provision? If so, what are they? What are the possible unintended negative consequences of the provision?

13F. Does the change in laws affected by the tax provision open new opportunities for unintended avenues for tax avoidance or enable tax evasion or fraud? How do you plan to minimize these potential risks?

Design for Success Tools—Tax Provisions (draft)

13G. How does the proposal affect the distribution of the overall tax burden?

C. Implementability

14. How will the program be implemented?

14A. Briefly describe the expected timeline for implementing the tax provision. Why do you think this timeline is realistic?

14B. Has the IRS been consulted on the tax provision? Does the IRS have the necessary systems and human resources to administer the program? What will the costs of administering the program be?

14C. If there is administrative discretion in allocating benefits (for example, awarding grants or authority) and you are proposing that the IRS should exercise that discretion, explain why you are not proposing that the provision be administered fully or in part by another agency. Is there another agency that is better suited due to subject matter expertise or access to information? Consider whether the IRS should administer the tax provision with policy decisions delegated to another agency.

14D. Are you confident that the appropriate aspects are detailed in Treasury Regulations rather than codified in the Internal Revenue Code?

14E. Will the tax provision require taxpayers to do anything to claim the benefit or comply with additional requirements? If so, describe how taxpayers will be provided with the information they need. For tax benefits, describe how you plan to maximize take-up among eligible taxpayers.

Design for Success Tools—Tax Provisions (draft)

14F. What additional information will need to be collected from taxpayers to administer the program? Explain what forms the IRS will need to develop or change to collect this information.

14G. Explain how the proposal will increase or decrease the time and expense of tax compliance for households and/or businesses.

15. How will you monitor success and rethink the approach?

15A. How will you know whether the tax provision is successful at achieving its objectives? List the key performance indicators for the provision.

15B. Will data that is collected provide a timely and accurate reflection of whether the tax provision is working? Will the data be comparable to that collected by other programs that have similar objectives (including spending programs)?

15C. Lay out the plan for evaluating whether the provision is successful (including when you expect the provision to be evaluated).

15D. What are the potential obstacles to phasing out or terminating the tax provision if it proves ineffective at achieving the desired outcome or is no longer deemed necessary?

Appendix 6

Reviewing What Works:
Tools for evaluating existing programs

Reviewing What Works—Policy Strategy Tool

Description

This tool is the first step in the Reviewing What Works process to assess the relative effectiveness of existing programs. An interagency panel uses this tool to look across a policy area (such as homelessness) to identify the key outcome goals (such as ending homelessness among veterans in five years). The panel then compiles an inventory of all federal programs that seek to achieve those goals. One policy strategy tool should be completed for each policy area. For each program on this inventory the lead agency completes the program effectiveness tool. Once all those tools have been completed the interagency panel then considers each program relative to others and completes a program evaluation tool for each program.

Policy area: _____   **Reference #:** _____

Interagency panel name: _____   **Date:** _____

A. Performance goals

1. What are the administration's priority goals in this policy area? Where appropriate, refer to high priority performance goals that have been adopted by agencies and are required by the GPRA Modernization Act of 2011. Include only goals that relate to what the program seeks to achieve for the American people—not expected outputs or expenditures. Set out for each goal the current level of performance, the targeted performance level, and the deadline. Do not list more than five goals across a policy area.

High priority performance goal/other goal #1: _____
Current level of performance: _____
Targeted performance level: _____
Deadline: _____

High priority performance goal/other goal #2: _____
Current level of performance: _____
Targeted performance level: _____
Deadline: _____

High priority performance goal/other goal #3: _____
Current level of performance: _____
Targeted performance level: _____
Deadline: _____

(List of goals continues as necessary to a maximum of five.)

B. Inventory of current programs

2. List the current programs that aim to achieve the goals above. For each program, provide the agency, bureau, description, and annual budget. List the programs in order of size of budget.

Program A name: _____
Agency: _____
Bureau: _____
Description: _____
Budget (for the current year and the previous three years): _____

Program B name: _____
Agency: _____
Bureau: _____
Description: _____
Budget (for the current year and the previous three years): _____

Program C name: _____
Agency: _____
Bureau: _____
Description: _____
Budget (for the current year and the previous three years): _____

Program D name: _____
Agency: _____
Bureau: _____
Description: _____
Budget (for the current year and the previous three years): _____

(List of programs continues as necessary.)

Reviewing What Works—Program Effectiveness Tool

Description

This tool is the second step in the Reviewing What Works process to assess the relative effectiveness of existing programs. Program managers use this tool to report on the effectiveness of their program and its contribution to common goals across a policy area (such as reducing homelessness). They should respond to the questions in clear, concise narrative (roughly 200 words per response) and attach additional details as appendices. One program effectiveness tool should be completed for each program on the policy area's policy strategy tool. An interagency panel on the specific policy area considers the information in this form to assess the program's effectiveness relative to other federal programs in that area.

Program: _____ Reference #: _____

Agency: _____ Bureau: _____

Author: _____ Approved by: _____ Date: _____

A. Program description and goals

1. Describe the program and when it was launched.

2. What have the program's annual costs for federal government been for the last four years?

3. Lay out specific policy outcome goals (such as ending homelessness among veterans in five years) the program is contributing to, referring to the information on the policy strategy tool.

4. Are there any other outcome goals the program seeks to accomplish? If so, list them briefly. Include only goals that relate to what the program is seeking to achieve for the American people. Do not include expected outputs or expenditures.

B. Program performance

5. Is the program working?

5A. Describe in brief how the program works. How does it seek to alter the ways of working or behavior of stakeholders or beneficiaries to achieve the outcome goals above?

5B. What evidence do you have that the program is effective at achieving the goals above? If possible, describe the precise contribution you believe that the program has made to the outcomes (such as reduced chronic homelessness by 2,500 across 17 states) and over what time period.

Reviewing What Works—Program Effectiveness Tool

5C. How confident are you that the contribution of the program mentioned in question 5B does not result from other programs or tax and regulatory measures? Why?

5D. Has there been an independent evaluation of the program and its effectiveness? If so, what was the methodology? In particular, was there a randomized control trial or similarly rigorous independent evaluation?

6. How does the program work alongside other programs?

6A. In addition to programs listed on the accompanying form, what other federal, state, local, and other programs seek to accomplish similar outcomes? What tax and regulatory measures seek to accomplish these outcomes?

6B. How does the program leverage or coordinate with other programs to minimize duplication and maximize collective impact?

7. What does benchmarking show about the costs and effectiveness of the program?

7A. Have you benchmarked the costs and effectiveness of the program with other programs that seek to achieve similar aims (including at the federal, local, or state level)? If so, what does the benchmarking show? How can you be sure that the comparison is fair and does not compare different types of beneficiaries or effectiveness (for example, short-term unemployed with long-term unemployed)?

7B. Have you benchmarked the costs and effectiveness of the program internally or across different geographical areas that the program operates in, across different groups the program works with, or using different program models? Do you have a strong understanding of the drivers of differences in cost and effectiveness? What have you done to ensure that the program maximizes learning from these cost comparisons?

8. Is the program well run?

8A. Has the program been implemented in line with the expected timeline? If not, describe the nature of delays and whether any delays are likely in the future.

Reviewing What Works—Program Effectiveness Tool

8B. Does the agency responsible for administering the program have the necessary staff, skill base, and IT infrastructure? Discuss any issues the program is experiencing in these areas and how you are dealing with them.

8C. Who is responsible for implementing the program at the state or local level? Are they fully committed to and working toward the program's goals?

8D. Does the program have an up-to-date risk management plan? How was it developed? And has it been good at predicting what might go wrong and how to mitigate risks?

8E. How accurate have the program's estimates of likely costs been over the last four years?

8F. Have there been any issues with obligating funds? Does the program display a year-end peak in spending?

8G. Are procurement costs incurred to date in line with expectations? Are the necessary contract staff and review structures in place to ensure that procurement costs are minimized and terms are closely monitored?

8H. How does the program seek to minimize fraud? What instances of fraud have occurred and what has been done to manage the risk in the future?

Reviewing What Works—Program Effectiveness Tool

9. Has the program sought to learn from experience?

9A. How do you monitor whether the program has any unintended consequences or whether there are ways to cheat the system? What evidence is there of unintended impacts or ways to cheat the system? What have you done in response to them?

9B. How does the program seek to improve? Is there a way for beneficiaries of the program or those who are involved in its administration to submit feedback and suggestions for improvement? How has the program been reformed to take account of this feedback?

9C. How does the program keep abreast of developments at the state level, internationally, or in the private sector that might help to identify ways the program could be improved? Has the program been reformed and improved in light of this learning?

9D. Is the data being collected an accurate and timely reflection of the program's performance toward achieving the outcomes? To the extent possible, are you coordinating data collection with other programs?

9E. Discuss the process for evaluating the program's progress toward achieving its goals. How often do performance assessments take place and to whom is progress reported? What improvements have been made to the program as a result of the performance assessment process?

Reviewing What Works—Program Evaluation Tool

Description

This tool is the third step in the Reviewing What Works process to assess the relative effectiveness of existing programs. An interagency panel uses this tool to evaluate the program's effectiveness compared to other programs in the same policy area (such as reducing homelessness). The panel evaluates each program listed on the policy strategy tool, probing individual program managers on the contents of the program effectiveness tool. The panel then assesses each key effectiveness factor on a scale of 1-10 (where 1 is very weak and 10 is very strong), and provides a brief explanation of each assessment. One program evaluation tool should be completed for each program on the policy area's policy strategy tool. The final program evaluation tool, the validated program effectiveness tool, and the completed policy strategy tool are submitted to the relevant agency, the OMB, and Congress for consideration during the annual budget and appropriations process.

Program: _____   Reference #: _____

Agency: _____   Bureau: _____

Policy area: _____   Interagency panel: _____   Date: _____

A. Program description and goals

**1. Describe the program and when it launched.**

2. What has the program's annual costs for federal government been for the last four years?

3. Lay out specific policy outcome goals (such as ending homelessness among veterans in five years) that the program is contributing to, referring to the information on the Policy Strategy tool.

4. Are there any other outcome goals that the program seeks to accomplish? If so, list them briefly. Include only goals that relate to what the program is seeking to achieve for the American people. Do not include expected outputs or expenditures.

B. Program performance

5. Is the program working?

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |

Very weak   Weak   Neither weak nor strong   Strong   Very strong
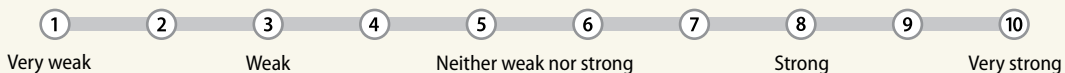
Weak program performance
Unsuccessfully seeks to alter ways of working or behavior of stakeholders or beneficiaries; no evidence that the program is effective, or that the program contributes to observed outcomes; no independent evaluation of the program's impact, or research shows no positive effect

Strong program performance
Successfully alters ways of working or behavior of stakeholders or beneficiaries; reliable evidence the program is contributing to observed outcomes; impact validated by rigorous independent evaluation

Reasons for evaluation

Reviewing What Works—Program Evaluation Tool

6. How does the program work alongside other programs?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak   Weak   Neither weak nor strong   Strong   Very strong
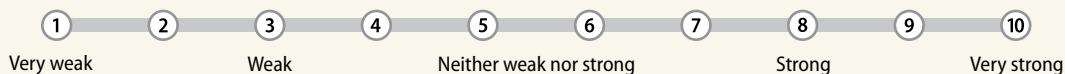
*Weak coordination with other programs*

High duplication of activities or services with other federal programs; program does not leverage or coordinate with similar programs

Strong coordination with other programs

Little or no duplication of activities or services with other federal programs; where overlap exists, program leverages and coordinates with similar programs whenever possible

Reasons for evaluation

7. What does benchmarking show about the costs and effectiveness of the program?

1 — 2 — 3 — 4 — 5 — 6 — 7 — 8 — 9 — 10

Very weak   Weak   Neither weak nor strong   Strong   Very strong
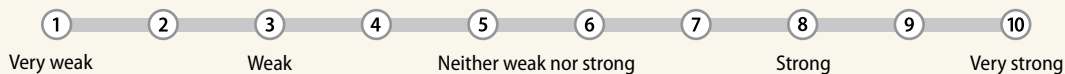
Weak cost effectiveness

No attempt to benchmark the costs and effectiveness of the program with other programs, or costs are excessive compared to similar programs with similar beneficiaries; no attempt to benchmark costs internally (across geographical areas, program groups, etc); weak understanding of drivers of cost and effectiveness, with no attempt to learn from comparisons
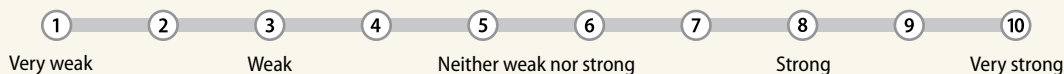
Strong cost effectiveness

Costs are reasonable relative to similar programs with similar beneficiaries; managers have a strong understanding of drivers of cost and effectiveness, and have taken actions to learn from comparisons

Reasons for evaluation

8. Is the program well run?

①—②—③—④—⑤—⑥—⑦—⑧—⑨—⑩

Very weak   Weak   Neither weak nor strong   Strong   Very strong

Weak program management

Significant delays in implementation; administering agency lacks necessary staff, skill base, or IT infrastructures, with insufficient plan to fix the problem; those in charge of implementing the program not committed to and working toward the program's goals; risk management plan either does not exist or is insufficient; actual costs much higher than initial cost estimates; significant issues with obligating funds; spending much higher near the end of the year relative to the rest of the year; procurement costs substantially higher than expected, or contract staff an review structures insufficient; high instance of fraud with insufficient strategy to mitigate these risks in the future

Strong program management

Implemented in line with the expected timeline; All necessary staff, skill base, and IT infrastructures are in place, or still on pace to attain them as expected; those in charge of implementing the program fully committed to and working toward the program's goals; strong risk management plan that has proven effective at predicting mitigating risks; actual costs less than or consistent with initial cost estimates; no issues obligating funds, with reasonably consistent spending throughout the year; procurement costs less than or consistent with expectations; necessary contract staff and review structures in place; low instance of fraud with a strong strategy to mitigate these risks in the future

Reasons for evaluation

9. Has the program sought to learn from experience?

①—②—③—④—⑤—⑥—⑦—⑧—⑨—⑩

Very weak   Weak   Neither weak nor strong   Strong   Very strong
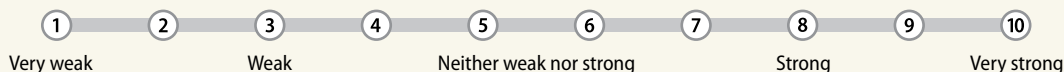
Weak learning process

No effort to monitor or respond to unintended impacts or ways to cheat the system; no way for beneficiaries or administrators to submit feedback and suggestions for improvement; managers do not keep abreast of relevant developments at the state level, internationally or in the private sector; insufficient data collected, or data do not reflect program performance toward achieving outcomes; no effort to coordinate data collection with similar programs; insufficient process for evaluating program progress

Strong learning process

Unintended impacts and ways to cheat the system closely monitored; effective steps have been taken to respond to unintended consequences and cheating; system in place for beneficiaries or administrators to submit feedback and suggestions for improvement, and managers take feedback into account; managers keep abreast of relevant developments at the state level, internationally or in the private sector, and improve the program accordingly; accurate and timely data being collected that reflects program performance toward achieving outcomes; coordinates data collection with similar programs where possible; effective process for evaluating program progress, and improvements are made based on the results

Reasons for evaluation

Endnotes

1   U.S. Census Bureau, "Census 2000 Completes Initial Work on All Known Housing Units; Field Operations Shift to 'Quality Counts' Phase," Press release, June 29, 2000, available at http://www.census.gov/newsroom/releases/archives/census_2000/cb00-cn47.html.

2   After completing the 2010 Census, the bureau actually ended under budget. In August 2010, the bureau returned $1.6 billion of its almost $7 billion operational budget. See "Census Bureau returns $1.6B of its operational budget," *CNN Money*, August 10, 2010, available at http://money.cnn.com/2010/08/10/news/economy/Census_Under_Budget.cnnw/index.htm.

3   Allan Holmes, "On the Brink: The success of the 2010 census hangs on a risky switch to handheld computers," *Government Executive*, July 15, 2007, available at http://www.govexec.com/features/0707-15/0707-15s2.htm.

4   Guy Molyneux, Ruy Teixeira, and John Whaley, "Better, Not Smaller, What Americans Want From Their Federal Government" (Washington: Center for American Progress, 2010), available at http://www.americanprogress.org/issues/2010/07/what_americans_want.html.

5   Pew Center on the States, "Facing Facts: Public attitudes and fiscal realities in five stressed states" (2010).

6   The SES is comprised of a dedicated corps of executives selected for their exceptional leadership abilities to transform the federal government and make it more citizen-focused and result-oriented. They act as the link between the top presidential appointees and the rest of the federal workforce. Partnering with Government Executive Magazine, William Eggers and his co-author John O'Leary reached out to the membership of the SES between March 2008 and April 2008 to get their input on what government can do to improve its execution ability. Most of the 217 respondents have more than 30 years of experience at the federal level. William D. Eggers and John O'Leary, *If We Can Put a Man on the Moon: Getting Big Results in Government* (Harvard Business School Press, 2009), Appendix B.

7   Ibid.

8   The agency's electric loan program provides low-interest loans to rural areas where electric utility costs take up a large portion of resident income.

9   President's Office of Management and Budget, "Terminations, Reductions, and Savings, Budget for the U.S. Government Fiscal Year 201" (2010), p. 30, available at http://www.gpoaccess.gov/usbudget/fy10/pdf/trs.pdf.

10   In 2002, Johns Hopkins professor Lester Salamon compiled a nearly 700-page anthology of political decision-making entitled *The Tools of Government: A Guide to New Governance*. The book outlines the many ways that governments—often in tandem with private and nonprofit entities—pursue solutions to public problems, what Salamon calls "tools for public action." Salamon argues that policymakers must constantly weigh their options in deciding not just whether, but also how, government should act when faced with pressing problems. He lays out 15 different tools for public action in the introduction to his book: direct government spending; government corporations; economic regulation; public information; direct loans; social regulation; contracting; loan guarantees; grants; tax expenditures; fees and charges; insurance; tort law; vouchers; and government-sponsored enterprises. Salamon argues that when making policy decisions, the central focus should be on which type of tool best serves the intended public purpose. This should be a technical decision, assessing the expected cost effectiveness, manageability, and basic fairness of each alternative. Lester Salamon, *The Tools of Government: A Guide to the New Governance* (New York: Oxford University Press, 2002).

11   United States Government Accountability Office, "Secure Border Initiative: DHS Needs to Reconsider Its Proposed Investment in Key Technology Program" (2010).

12   Chuck McCutcheon, "Securing Our Borders Doing What Works to Ensure Immigration Reform Is Complete and Comprehensive" (Washington: Center For American Progress, 2010), available at http://www.americanprogress.org/issues/2010/04/secure_borders.html.

13   Interview with former DHS official, January 2011.

14   David A. Powner and Mathew J. Scirè, "Information Technology, Significant Problems of Critical Automation Program Contribute to Risks Facing 2010 Census," Testimony before Senate Committee on Homeland Security and Government Affairs, March 5, 2008.

15   Chris Strohm, "DHS Terminates Virtual Border Fence Program," *National Journal*, January 14, 2011, available at http://nationaljournal.com/nationalsecurity/dhs-terminates-virtual-border-fence-program-20110114.

16   Holmes, "On the Brink."

17   Ibid.

18   United States Government Accountability Office, "Information Technology: Significant Problems of Critical Automation Program Contribute to Risks Facing 2010 Census" (2008), available at http://www.gao.gov/highlights/d08550thigh.pdf.

19   David A. Powner and Mathew J. Scirè "Information Technology: Census Bureau Needs to Improve Its Risk Management of Decennial Systems," Testimony before the Subcommittee on Information Policy, Census, and National Archives, Committee on Oversight and Government Reform, U.S. House of Representatives, December 11, 2007, available at http://www.gao.gov/new.items/d08259t.pdf.

20   Dana Hedgpeth, "Census Back to Pen and Paper: Bureau Admits It Wasn't Clear on Computer Needs," *The Washington Post*, April 4, 2008, available at http://www.washingtonpost.com/wp-dyn/content/article/2008/04/03/AR2008040302068.html.

21   Gautham Nagesh, "Mismanagement, not technology, caused Census handheld trouble, auditors say," *Government Executive*, April 9, 2008, available at http://www.govexec.com/dailyfed/0408/040908n1.htm.

22   Peter T. Kilborn, "Law Fails to Stem Abuse of Migrants, U.S. Panel Reports," *New York Times*, October 22, 1992, p. A1, available at http://www.nytimes.com/1992/10/22/us/law-fails-to-stem-abuse-of-migrants-us-panel-reports.html?ref=immigrationandemigration.

23   Another weakness of the IRCA was that it failed to create legal channels for those seeking future employment in the United States and the outdated family visa ceilings.

24 According to a Pew report, the total population of undocumented immigrant population grew rapidly from 1990 to 2006 but has since stabilized. Jeffrey S. Passel and D'Vera Cohn, "A Portrait of Unauthorized Immigrants in the United States" (Washington: Pew Hispanic Center, 2009), available at http://pewresearch.org/pubs/1190/portrait-unauthorized-immigrants-states.

25 "Arne Duncan 'No child' is 'toxic'" On Point Radio, March 5, 2009, available at http://www.onpointradio.org/2009/03/arne-duncan-no-child-is-toxic.

26 The system of incentives and penalties established in the law created powerful motivations for schools, districts, and states to manipulate test results. In one state, education officials lowered the passing score on the eighth-grade math test. Other states simply made their tests easier and easier each year. See, Eggers and O'Leary, *If We Can Put a Man on the Moon*, p. 2.

27 According to the Department of Education's National Assessment of Education Progress, just 26 percent of twelfth-graders performed at or above the "proficient" level in math in 2009. Thirty-eight percent performed at or above the "proficient" level in reading that year. Both percentages are well below NCLB's target of 100 percent reading and math proficiency across the country by 2014. National Center for Education Statistics, "The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results" (2010), available at http://nces.ed.gov/nationsreportcard/pubs/main2009/2011455.asp.

28 Atul Gawande, *The Checklist Manifesto: How to Get Things Right* (New York: Metropolitan Books, 2009), p.185.

29 For a summary of key questions that the Regulatory Impact Analysis covers, see the "Agency Checklist: Regulatory Impact Analysis" (Office of Information and Regulatory Affairs, available at http://www.whitehouse.gov/sites/default/files/omb/inforeg/regpol/RIA_Checklist.pdf).

30 Peter Orszag, "Remarks on Government Reform and Fiscal Responsibility" (Washington: Center for American Progress, 2010), available at http://www.americanprogress.org/events/2010/06/inf/Orzag-transcript.pdf.

31 The National Commission on Fiscal Responsibility and Reform, "The Moment of Truth" (2010), available at http://www.fiscalcommission.gov/sites/fiscalcommission.gov/files/documents/TheMomentofTruth12_1_2010.pdf.

32 Adam Hughs, "Changes to PART Referenced in Obama Budget," *OMB Watch*, May 11, 2009, available at http://www.ombwatch.org/node/9982.

33 United States Interagency Council on Homelessness, *Opening Doors, Federal Strategic Plan to Prevent and End Homelessness 2010*, http://www.usich.gov/PDF/OpeningDoors_2010_FSPPreventEndHomeless.pdf

34 Although many new program ideas do come out of the executive branch, an agency can normally only create a new program if it has legislative authority to do so. Even if the program is "created" in the agency, they are normally grounded on an act of Congress (known as "authorizing legislation"). There are some instances where the executive branch lobbies Congress to create new authorizing legislation—we propose that the legislative process should be followed in those cases.

35 According to the Congressional Research Service, roughly two-thirds of federal spending for agencies and programs is provided by authorizing legislation (known as direct spending). Some direct spending—mostly in the form of entitlement programs—is funded by permanent appropriations in the authorizing law. The remaining one-third of funding is determined through the annual authorization and appropriations process (known as discretionary spending). For the most part, agencies request funding for discretionary programs through an annual submission to the Office of Management and Budget, which informs the president's budget. Bill Heniff Jr., "Overview of the Authorization-Appropriations Process" (Washington: Congressional Research Service, 2010), available at http://www.senate.gov/CRSReports/crs-publish.cfm?pid=%270DP%2BPLW%3C%22%40%20%20%0A.

36 There are a few instances where the review process is more formalized within the executive branch, but none deal directly with new spending programs. For new capital projects, the OMB-300 sets questions that agencies must answer and pertain to a new projects likely success—but capital projects are only a small part of the new initiatives that are considered every year. For new rules and regulations, each agency is required to complete an impact assessment, laying out the likely cost to business and others, and the likely benefits from the proposed rule. But new spending programs are not subject to similar formalized scrutiny.

37 Often appropriators devise new program ideas that are not in the president's proposed budget. In such cases, we believe they should use the legislative tools to test those ideas for likely success. By this stage in the process, the information in both the details and assessment tools will have been validated by an independent, interagency panel. So appropriators should be able to draw some comfort on the accuracy of the ratings.

38 We are aware that the actual political process is often much more complex than the framework we've presented. Each new government program has to pass through a series of internal bureaucratic hoops. Each new bill can be fast-tracked or blocked by countless congressional rules. What we're presenting here is one possible solution based on a simplified model —a recommendation to ensure the essential questions are being asked of every new government initiative.

39 As we discuss in the main report, tax measures are somewhat different but do require a similar level of scrutiny for likely success. This report includes a draft program details tool that might be used to evaluate proposed tax measures. The legislative version of the description tool should be a bit different from the executive tool. Often new programs created by legislation are still budding ideas, without concrete details of how the program will be carried out. To be sure, any new program should be able to answer basic questions about incentive structures, evidence of effectiveness, and implementability. But we expect the amount of detail in those answers to be less on Capitol Hill than in the agencies.

40 Again, we are not proposing that Congress conducts a full Design for Success review for every program created by a piece of legislation. As bills often establish small initiatives and pilot programs, we recommend only reviewing initiatives above a certain cost threshold (say $50 million).

41 This initial draft of the tax tool was developed by Seth Hanlon and Jitinder Kohli at the Center for American Progress. One of the co-authors of the study, Wiliam Eggers, works for Deloitte and due to the nature of their work precluded himself from the development of this tool.

Jitinder Kohli is a Senior Fellow on the Doing What Works project at American Progress. His work focuses on government efficiency, regulatory reform, and economic issues. Prior to joining CAP, Kohli spent 15 years in the British government and led the design and implementation of a number of major programs. He was the first chief executive of the Better Regulation Executive from 2005 to 2009, where he was responsible for regulatory reform and led efforts to reduce the administrative burden of regulation by 25 percent. He has also served as the director general of strategy and communications for the Department for Business, Innovation and Skills and as head of productivity and structural reform in Her Majesty's Treasury. He has also served as director, Active Communities and head of the Community Cohesion Unit in the Home Office. Prior to that, he worked in the Prime Minister's Strategy Unit.

William D. Eggers is a director for Deloitte Research,* part of Deloitte LP, and is the research director of Deloitte's GovLab and its Public Leadership Institute. Eggers is a former appointee to the U.S. Office of Management and Budget's Performance Measurement Advisory Commission and the former project director for the Texas Performance Review/e-Texas initiative. Eggers also served as a commissioner for the Texas Incentive and Productivity Commission and a designee on the Texas Council on Competitive Government. He is a former senior fellow at the Manhattan Institute for Policy Research and the former director of Government Reform at the Reason Public Policy Institute, a Los Angeles-based think tank. Eggers is the author of numerous books on government reform, including *If We Can Put a Man on the Moon… Getting Big Things Done in Government* (Harvard Business Press, 2009), *Letting Go of the Status Quo: A Playbook for Transforming State Government* (Deloitte 2010), *The Public Innovator's Playbook* (Deloitte Research 2008), *Government 2.0* (2005), *and Governing by Network* (2004).

John Griffith is a Research Associate with the Doing What Works project at American Progress. His work focuses on government efficiency and budget analysis. Prior to joining CAP, Griffith worked as an analyst with the Social and Economic Policy Division of Abt Associates, where he conducted research on federal homeless assistance, public housing, and community development programs. He previously interned with the United Nations Environment Program in Geneva, Switzerland, and the Foreign Policy Research Institute in Philadelphia.

Acknowledgements

This project was a collaborative effort. We'd like to start by thanking the staff at the Center for American Progress for their assistance throughout the project. We are extremely grateful to Lauren Bazel, Donna Cooper, Karen Davenport, Gadi Dechter, James Hairston, Seth Hanlon, Reece Rushing, Shannon Ryan, Raj Sharma, Lauren Smith, Isha Vij, and Sarah Wartell at CAP provided ongoing feedback, guidance, and editorial support for the project.

Our tools benefited from the input of literally hundreds. We are particularly indebted to those who helped us organize workshops and would like to thank Tiffany Dovey from Deloitte, John Kamensky and Jonathan Breul from the IBM Center for the Business of Government, Kristine Lam with the Senate Committee of Homeland Security and Government Affairs, Kathy Stack and Mary Cassel at OMB, Bernice Steinhardt and Adam Miles at GAO, Bethany Blakey at the Department of Transportation, Celeste Ritchie from HHS, and Ted Dallas at the Office of Maryland Governor Martin O'Malley. We are grateful for the insights and feedback from all who participated in the workshops, including the dozens of attendees at the July 2010 Doing What Works Conference in Washington, D.C.

In addition, we would like to extend special thanks to Jon Baron, Evie Barry, Clint Brass, Dustin Brown, Carl Davis, Amy Edwards, Michael Filler, Harry Hatry, Don Kettl, Roy Kienitz, Bob Knisely, Jeffrey Liebman, Nat Logar, Anthony Love, Shelley Metzenbaum, Kathryn Newcomer, John Palguta, Paul Posner, Sharon Price, Steve Redburn, Shep Ryen, Lester Salamon, Robert Shea, and Russ Whitehurst for their invaluable advice and feedback.

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is "of the people, by the people, and for the people."

Center for American Progress