

What leads to a successful evaluation? Reflections from 100 Office of Evaluation Sciences collaborations

OES has completed 100 collaborations across government since 2015. This has entailed working with agency collaborators across 21 federal departments, in addition to 139 OES team members from 83 institutions. Our evaluations have involved over 44 million individuals. In this document, we share our reflections on what we have learned about how to build and use evidence in government to date across 100 completed collaborations.

Reflection 1: A program change and evaluation approach don't have to be ideal to be valuable.

Launching a successful evaluation involves creative problem solving, flexibility, and a deep understanding of how a program or program changes may impact people. The ideal program change and approach to evaluation may be different than what is often feasible to do in a government or real world context. It can be possible to build evidence and improve government programs even when the ideal intervention or evaluation approach is not feasible. The collaboration below allowed us to field a promising intervention and build evidence using rigorous, if not "ideal," evaluation methods.

OES Example

[We collaborated with the U.S. Department of Agriculture](#) (USDA) to increase responses to income verification requests in the National School Lunch Program, and designed a new request letter that used simpler language, new visuals, and a personalized message from Local Education Agencies (LEA). An ideal evaluation strategy may have been to randomize which LEAs used the new letters, but this was not possible. Instead, USDA and OES worked with 74 LEAs to implement the changes with redesigned verification letters and compared verification responses in these LEAs to 82 randomly selected LEAs that were not asked to revise their letters. We learned that improving the design and clarity may be helpful in increasing response to verification requests - results showed an imprecisely estimated 2.1% improved response rate - but that more substantial changes to the verification procedure will be necessary to increase response rates appreciably.

Reflection 2: Coupling randomized evaluations with administrative data let's us learn things quickly, rigorously and at low-cost.

Randomized evaluation using administrative data can answer meaningful questions, without requiring additional data collection. Administrative data has made it possible for OES to improve government programs with a collective sample of over 44 million people since 2015. In many cases administrative data can more accurately measure behaviors and actions than self reported responses via a survey. Agencies often have underutilized rich sources of administrative data, including short-term outcomes such as online engagement data such as open and click rates, intermediate outcomes such as program take-up, and downstream effects such as college enrollment. Rigorous evaluations can be implemented and analyzed in days, weeks or months, without additional data collection costs or delays.

OES Example

The federal financial aid system plays a key role in expanding postsecondary access and affordability for students struggling to pay for college. However, many students that could benefit most from federal financial aid struggle to complete the Free Application for Federal Student Aid (FAFSA), leaving an estimated \$2.6 billion in federal student aid on the table each year. In collaboration with four federal agencies, [OES has built a](#)

[portfolio of evidence](#) on the impact of low-cost interventions that are designed to improve access to federal financial aid to help pay for college, postsecondary access and success, and student loan repayment. This portfolio leverages agency's administrative datasets and utilized data sharing agreements across agencies to rigorously evaluate priority questions.

Reflection 3: There's more to sample size than counting the number of participants in a program. The number of participants in a program does not always provide a good indication of an ability to detect impact, because there's more to statistical precision than sample size and there's more to sample size than counting. Anticipating the many ways that sample sizes grow and shrink can be instrumental to successful evaluation, and administrative data may provide data for multiple time points and numerous observations to offset these challenges.

OES Example

[We collaborated with the Small Business Administration](#) and a city government to study a lottery that distributed grants and loans to small businesses in the wake of the COVID-19 pandemic. While over 1,400 businesses applied for the grants and loans and were initially entered into the lottery, the number of businesses in our final sample shrank dramatically, due to businesses entering both the grant and the loan lotteries and a low match rate between the business dataset and the outcome dataset we used. We were able to offset these losses somewhat because we could gather administrative data at multiple time points.

Reflection 4: Being transparent about implementation — as well as results — has enormous value. Our commitment to transparency allows us to ensure agencies are benefiting from the full set of results, helps us fill evidence gaps about what works and what does not, and helps us build trust inside and outside of government. Documenting and sharing the details of an intervention or program that was evaluated, including the audience and how the intervention was delivered, is an important and often overlooked step.

OES Example

We collaborated with [the U.S. Agency for International Development](#) and Population Services International to increase the effectiveness of clinic referrals for family planning services using evidence-based insights. We designed a series of eight messages to send via text to women, reminding them of their referral for family planning and encouraging them to visit a clinic. In our commitment to transparency, prior to looking at outcome data, we posted an [Analysis Plan](#). During the evaluation, we carefully monitored the implementation and tracked enrollment over time. After the evaluation was completed, we posted an [intervention pack](#) to our website, to allow our audiences to interact with what we did, as well as to potentially incorporate elements into their own work. By sharing details of interventions, we ensure federal collaborators and the public can learn from—and reproduce—what we do.

As agencies build their evaluation capacity, we hope our lessons learned and resources can be of use. We have details on our [project process online](#), including templates for general use. We have also recently completed a [behavioral insights guide to improving payment integrity](#), incorporating lessons learned from the OES portfolio that we hope can be applied to agency challenges. We continually update [methodological guidance](#) which we have developed and found useful as we have conducted these evaluations. We have [summarized information about bringing external talent into government](#), as well as information on [all OES team members to date](#).