

OES Evaluation Policy

At the Office of Evaluation Sciences (OES), because we generate results that impact the lives of millions of Americans, the quality of our work and reliability of our findings are of paramount importance. We take this responsibility very seriously, and we have developed an Evaluation Policy designed to ensure that our evaluations are conducted to the highest standards.

This evaluation policy statement presents core principles that guide OES' planning, implementation, and use of testing, or evaluation, to learn what works. The way in which OES conducts evaluations should always achieve the following tenets:

- **Rigor** — Our findings should be credible and mean what they purport to mean. Evaluations are conducted to the highest standards; our statements about statistical significance are clear and correct; and the limitations of our findings are clear. All OES analyses go through an internal replication, and any discrepancies between the two analyses are addressed before the results are finalized. Bottom line: Policymakers and program designers should be able to act on our findings with confidence.
- **Relevance** — Project selection should take into account (1) the policy or program priorities of our agency collaborators, and (2) potential impact on a priority outcome based. To OES, relevance is also ensured by exclusively utilizing existing administrative data as the outcome of our evaluations.
- **Transparency** — We are committed to ensuring that researchers, agency collaborators, policymakers, and the public at large are able to learn from our work. We keep a public record of all evaluations fielded and publicize all of our findings (including null results and those that run counter to our own prior expectations and goals). We conduct our work in such a way that we can verify our results, and we document our work thoroughly so that others with required data access could reproduce the results.
- **Independence** — We retain control over decisions about project selection and the ability to release our evaluation results.
- **Ethical Practice** — OES evaluations will be conducted in an ethical manner and safeguard the dignity, rights, safety, and privacy of participants. Evaluations will comply with both the spirit and the letter of relevant requirements such as regulations governing research involving human subjects.

OES Testing Process

To implement our Evaluation Policy and ensure that we adhere to our core principles, we run all of our tests through a process that includes six **gates**. At each gate, the test is vetted against certain criteria before it can enter its next phase. Some of these gates emphasize project management considerations such as feasibility, planning, and clear documentation, while others emphasize specific methodological issues such as statistical power, pre-commitment to analysis plans, and reproducibility. The six gates are:

- I. **Partner with a Federal Agency Gate** - Each test is vetted early for feasibility, proper planning, and potential impact for stakeholders in a Federal program or policy.
- II. **Translate Evidence-Based Insights Gate** - Before any test can progress to the “field,” phase, its design is carefully reviewed by select team members and then presented to the full team for comments. Additionally, before we use code for random assignment, we make sure it has been independently reviewed by an OES team member who is not directly involved in the project.
- III. **Analysis Plan Commitment Gate** - Before we work with data, we commit to an Analysis Plan specifying key outcomes and analytic choices, and we post it on our website.
- IV. **Embeds Tests using Randomized Evaluation Gate** - An initial analysis of results is presented to the full team to check that the analysis is sound and comprehensive, that any limitations have been identified, and that alternative explanations have been addressed to the greatest extent possible.
- V. **Ensure Our Work Meets Evaluation Best Practice Gate** - After the initial analysis has been team-vetted and refined, it is checked by having a second analyst, who is unaware of the initial results and findings, reproduce it. Any discrepancies between the two analyses are addressed before the results are finalized.
- VI. **Measure Impact and Build Evidence Gate** - To ensure transparency and reproducibility, all test materials are checked for completeness and proper archiving before a report is published.

