

Introduction

Background.

The *Foundations for Evidence-Based Policymaking Act of 2018* (“Evidence Act”) requires agencies to develop a multi-year learning agenda that is used to identify agency priority questions. It further requires agencies to develop an Annual Evaluation Plan that provides a more detailed description of **significant program evaluations** that the agency plans to conduct in the following fiscal year. Agencies can define “significant” for this purpose.

Approach.

The Treasury FY 2022 Annual Evaluation Plan is the agency’s first evaluation plan, so it also includes activities planned for FY 2021. Treasury is currently developing an Evaluation Policy that will establish ways for the Evaluation Officer to better track planned significant evaluations, methodologies, and dissemination of findings. Treasury is also developing an assessment of Treasury’s capacity for evidence building (“Capacity Assessment”), which will be used to identify additional opportunities for evidence-building activities, including evaluations, across the agency. For the purposes of the draft Capacity Assessment and Evaluation Plan, the four types of evidence-building activities are defined as follows (subject to refinement as policies are further developed):

Evaluation	Collection and analysis of data to assess effectiveness and efficiency of programs, policies, or procedures
Statistics	Collection, compilation, and processing of data for describing or estimating characteristics or insights concerning groups
Research	Modeling or other systematic use of data to explore emerging issues or potential scenarios to generate new knowledge
Analysis	Routine and frequent use of data that produces insights for decision making and program management

Treasury has also developed a working definition for the term “significant evaluation” (pending approval of a final Evaluation Policy): *evaluations that are undertaken to evaluate critical programs, as defined in the Treasury implementation plan for the Program Management Improvement Accountability Act (PMIAA), and for which the evidence generated by the evaluation is relevant to a priority research question in the Treasury Learning Agenda.* Evaluations required by law are also considered significant.

The Treasury PMIAA implementation plan defines critical programs as programs that are essential to Treasury’s successful execution of its mission. Failure, interruption, or compromise of these programs could:

- 1) Undermine Treasury’s capability to achieve its objectives; or,
- 2) Severely impact government operations; or,
- 3) Result in significant loss of trust in the Department by key stakeholders/the public.

Dissemination of results for Treasury Evaluation Plan Projects

With the publication and implementation of the agency evaluation policy in FY 2022, Treasury plans to develop a standard for the dissemination of evaluation results and findings. Over the coming year, with input from the bureaus and offices that regularly conduct evaluations, the Evaluation Officer, Chief Data Officer, and Statistical Official will determine the appropriate mechanism for dissemination of findings internally and externally. Future evaluation plans will include more specific dissemination plans for each evaluation.

Evaluation Plan Contents

Overview of Planned FY 21 & 22 Evaluation Projects	2
Details associated with Planned Evaluation Projects	3
Treasury Strategic Objective: Tax Policy and Administration	3
Question 1.....	3
Question 2.....	4-5
Treasury Strategic Objective: Domestic Recovery and Growth	6
Question 3	6

Treasury FY 2022 Annual Evaluation Plan

Overview of Planned FY 21 & 22 Evaluation Projects

Treasury plans to conduct five significant evaluation projects, all starting in FY 2021, aligned to three learning agenda priority research questions. All of these evaluations align to the planned FY 2022 – 2026 Strategic Goal “Promote Equitable Economic Growth and Recovery.”

1. How can the IRS and TTB use Artificial Intelligence and other techniques to improve examination, fraud and criminal case selection methods, algorithms and other processes to improve compliance and diminish repeat noncompliance?
 - a. **Project 1.1:** Marginal Revenue/Cost Case Selection for Campus Exam
2. How can the IRS address taxpayer preferences and needs through a comprehensive taxpayer experience strategy, in line with the requirements of the Taxpayer First Act?
 - a. **Project 2.1:** Measuring the Impact of Proactive Outreach and Education
 - b. **Project 2.2:** Phase II Usability Testing of Online Account
 - c. **Project 2.3:** Customer Callback initiative
3. To what extent did the Paycheck Protection Program forestall further unemployment and bankruptcies during the pandemic?
 - a. **Project 3.1:** Economic Modeling of recovery program effects with MITRE

Additionally, Treasury will continue to explore conducting significant evaluation activities for four other learning agenda priority research questions. These activities may start by the end of FY 2022, dependent on resources:

Strategic Goal	Priority Question	Lead organization
Promote Equitable Economic Growth and Recovery	What steps can the IRS/TTB take to prioritize and address emerging noncompliance issues?	IRS, TTB
Promote Equitable Economic Growth and Recovery	What are the impacts and/or outcomes of the Emergency Capital Investment Program, Rapid Response Program, and Minority Lending Programs on Community Financial Development Institutions (CDFIs) or minority depository institutions (MDIs)?	Departmental Offices
Enhance National Security	How can Treasury strengthen its situational awareness of the financial sector and encourage greater industry participation in the Office of Cybersecurity and Critical Infrastructure Protection (OCCIP) initiatives?	Departmental Offices
Modernize Treasury Operations	How are consumers using USAspending.gov?	Fiscal
Modernize Treasury Operations	To what extent would tailored assessment tools result in improved hiring outcomes (i.e. time to hire, quality of hire)? <i>Note: this question may be modified based on ongoing discussions with OPM and other agencies and efforts to craft a government-wide set of workforce questions</i>	Departmental Offices

Treasury FY 2022 Annual Evaluation Plan

Details associated with Planned Evaluation Projects

Title I of the Evidence Act requires that agencies share specific information for selected significant evaluation projects to help anticipate the completion of and plan for the appropriate dissemination of the findings for use in agency decision-making. The following tables (one for each project) provide the required information: the project title, lead research organization, supporting organization(s), specific evaluation questions for each project, data(set)s required for the project, the tools and software required for the project, the analytic approaches to be employed for the project, and anticipated/existing challenges to completing the project.

Priority Question 1: How can the IRS use Artificial Intelligence and other techniques to improve examination, fraud and criminal case selection methods, algorithms, and other processes to improve compliance and diminish repeat noncompliance?			
Project 1.1: Marginal Revenue/Cost Case Selection for Campus Exam Lead Organization: RAAS Supporting Organization(s): SB/SE			Project starting in FY21
Evaluation Questions:	Data	Tools	Analytic Approaches
Question 1: How can machine learning techniques improve case selection?	Successful execution of the project requires access to the following data sources on CDW: <ul style="list-style-type: none"> • AIMS_CLOSED • ERIS • IRTF • IRMF • NRP Individual Income Tax Other data obtained separately: <ul style="list-style-type: none"> • Unit cost data from CFO • Definitions of status quo selection filters and prioritization metrics 	<ul style="list-style-type: none"> • The entire range of data and tools available on CDW, in particular the R statistical software package, python, and boost • A prototype tax calculator we created • A tool we created (Model Explorer) to enable operational staff to apply the latest case selection models to prioritize new potential Examination inventory 	<ul style="list-style-type: none"> • “Age” historical audit data to current tax law • Use a wide range of variables in several competing machine learning approaches to predict revenue/cost in historical year(s) • Test models in later historical year(s) to identify Best Available Model (BAM) • Test BAM vs. status quo in operational random control trials
Question 2: How can estimating marginal revenue/cost improve case selection?			
Question 3 How can we be confident that new selection models are actually better than the status quo selection methods?			
Challenges associated with project:			
<ul style="list-style-type: none"> • Status quo selection methods have often changed during the years for which we have historical audit results • Each machine learning technique has pros and cons; need to have them compete against each other using common objectives and criteria • Getting complete audit results (e.g., revenue) often takes years after the exam closes, so we need to balance completeness of data against how recent the exam was • Some cost data are not available at the case level, so we may need to approximate such cost components using what is available • Changes and problems in computing environment can drastically stall progress • Complex data extraction workflows can be difficult to manage • Operational test results take years to complete due to the exam/appeals/collection process , but initial results are expected in FY22 			

Treasury FY 2022 Annual Evaluation Plan

Priority Question 2: How can the IRS address taxpayer preferences and needs through a comprehensive taxpayer strategy, in line with the requirements of the Taxpayer First Act?			
Project 2.1: Measuring the Impact of Proactive Outreach and Education Lead Organization: Taxpayer First Act Office Supporting Organization(s): W&I, SB/SE, C&L			Project starting in FY21
Evaluation Questions:	Data	Tools	Analytic Approaches
Question 1: How can the IRS measure the effectiveness of Proactive Outreach and Education materials?	<ul style="list-style-type: none"> • Proactive Outreach and Educational Materials provided to taxpayers • CDW (Compliance Data Warehouse) data • Taxpayer questions and inquiries • Work, Planning and Control Systems (WP&C) data • Other operational data sets that are not in CDW 	<ul style="list-style-type: none"> • Tableau • Microsoft Excel • SAS, SPSS (statistics packages) • Social media analytics, as appropriate • Tools such as CyberTrans could possibly be leveraged for translation services in communications 	<ul style="list-style-type: none"> • Test and Learn pilot tests • Quantitative data analytics (e.g., statistical studies) • Behavioral insights (e.g., profiles, journey analytics)
Question 2: How likely are taxpayers to contact the IRS after receiving targeted Proactive Outreach and Education materials?			
Question 3: To what extent are taxpayers more likely to accurately meet their filing requirements after receiving proactive outreach and education materials from the IRS?			
Challenges associated with project: <ul style="list-style-type: none"> • There is no centralized area/organization within the IRS for addressing and measuring the effectiveness of Proactive Outreach and Education materials • Gathering feedback from targeted taxpayers is cumbersome, time-consuming, and is often fraught with unreliable information 			
Project 2.2: Phase II Usability Testing of Online Account (OLA) Lead Organization: Wage & Investment Supporting Organization(s): N/A			Project starting in FY21
Evaluation Questions:	Data	Tools	Analytic Approaches
Question 1: How can taxpayer feedback be incorporated to tailor service and process design?	<ul style="list-style-type: none"> • Findings from iterative usability and user experience testing sessions • Payment history/activity page • Payment options- Installment agreement • E-authentication 	<ul style="list-style-type: none"> • One on one interviews with users • Focus Groups 	<ul style="list-style-type: none"> • Qualitative and quantitative analysis of testing session data, identifying key findings and thematic focus for subsequent testing rounds • Usability testing Likert Scale analysis • Theme coding of qualitative data-open ended questions
Question 2: Which OLA functionalities will help facilitate taxpayer adoption of OLA?			
Question 3: How can IRS address the identified user experience issue(s)?			
Challenges associated with project: Contractor support needed to recruit participants, secure facilities, and dispense incentives and take notes during sessions			

Priority Question 2 (cont.): How can the IRS address taxpayer preferences and expectations through a comprehensive taxpayer experience strategy, in line with the requirements of the Taxpayer First Act?

Project 1.3: Customer Callback Initiative Lead Organization: Joint Operations Center Supporting Organization(s): Compliance, Customer Account Services, WI Operations & Support	Project starting in FY21
---	---------------------------------

Evaluation Questions:	Data	Tools	Analytic Approaches
Question 1: How can IRS maximize the use of the latest contact center technology to improve taxpayer experience?	Evaluative Data for CCB Design includes: <ul style="list-style-type: none"> • Application Demand • Average Speed of Answer • Secondary Abandon Rate 	Management Information System reports include the following platforms with varying report intervals and stakeholder recipients: <ul style="list-style-type: none"> • Intelligent Contact Management (ICM) • Contact Analytics • Enterprise Telephone Data (ETD) • 3rd Party Tools - EyeQueue[®] Callback Dashboard & Portal 	<ul style="list-style-type: none"> • Measuring progress of Customer Callback applications deployed against IRS Mod Plan reportable measures. • Industry benchmarks for KPIs • Monitoring performance trends • Evaluating customer feedback through sampling of calls identified within CCB services
Question 2: How does Customer Callback reduce the amount of time taxpayers physically wait on hold/taxpayer burden?	<ul style="list-style-type: none"> • Secondary Abandon Time (seconds) • CCB Success Metrics (Primary) 		
Question 3: How effective is Customer Callback at facilitating quick and efficient resolution of taxpayer issues?	<ul style="list-style-type: none"> • CCB Offer Rate • CCB Take Rate • CCB Reacquisition Rate (to agent) • CCB Cancellation Rate • CCB Busy/No Answer Rate • Avg Reconnect Time (to agent) • Hold Time Saved • Customer Feedback via call recordings • CCB Success Metrics (Secondary) • Successfully Registered • Transfer Error 		

Challenges associated with project:

- IT Service Provider has limited implementation bandwidth against competing priorities.
- CCB Design is evolving with, and in parallel to, infrastructure upgrades.
- Customer callback offerings are limited by both system and human resources (agents).

Priority Question 3: To what extent did the Paycheck Protection Program forestall further unemployment and bankruptcies during the pandemic?			
Project 3.1: Economic modeling with MITRE Lead Organization: Economic Policy (Departmental Offices) Supporting Organization(s): Tax Policy (Departmental Offices), MITRE			Project starting in FY21
Evaluation Questions:	Data	Tools	Analytic Approaches
<p>Question 1: What were the net employment effects of PPP? How did those impacts change through the several rounds of revisions to the program?</p> <p>Question 2: What impact did PPP have on the level of bankruptcies? How did those impacts change through the several rounds of revisions to the program?</p>	<ul style="list-style-type: none"> • PPP loan data (SBA) • Employment data (BLS) • Census demographic and income data, business data • Bankruptcy data (DOJ) 	<ul style="list-style-type: none"> • Statistical packages 	<ul style="list-style-type: none"> • Quantitative data analytics (e.g., statistical studies using loan data matched to employment data) • Economic Modeling
<p>Challenges associated with project:</p> <ul style="list-style-type: none"> • Data availability and access (e.g., restricted-use data requiring interagency agreements and use at off-site facilities, missing or unreported data) • Data sensitivity (e.g., personally identifiable information) • Outside researchers addressing main results more quickly • Operational considerations (e.g., cost benefit analysis, limited resources) 			