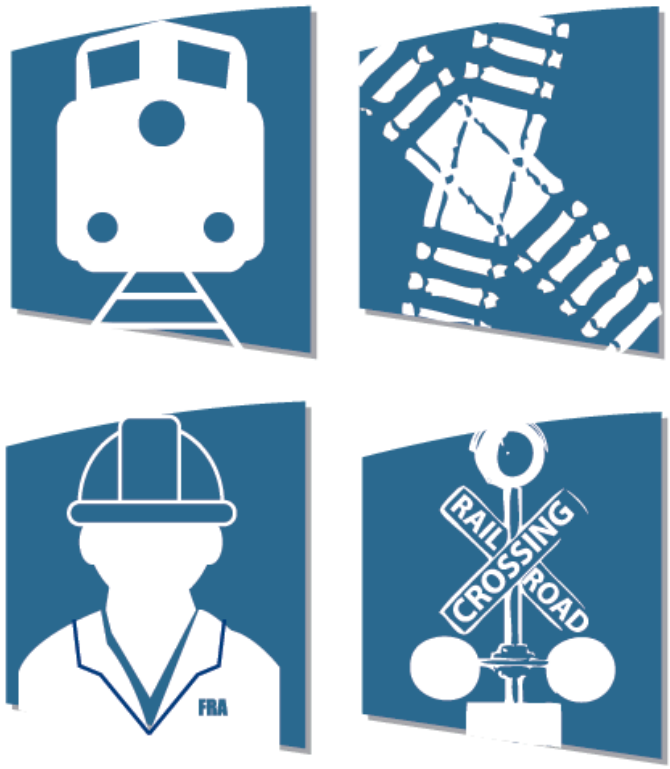
U.S. Department
of Transportation

Federal Railroad Administration

EVALUATION IMPLEMENTATION PLAN

Office of Research & Development



OFFICE OF RESEARCH & DEVELOPMENT

FRA OFFICE OF RAILROAD POLICY & DEVELOPMENT

EVALUATION VISION

After five years, the Office of R&D will excel in regularly employing sound evaluation to help improve the development, utilization, impact, and overall effectiveness of R&D products, tools, and processes.

EVALUATION MISSION

To embed program evaluation methods in R&D programs through education, sharing best practices and exemplar studies.

DOT/FRA/ORD-13/47

NOVEMBER 2013

NOTICE

The United States Government does not endorse products or manufacturers. Trade or manufacturers' names appear herein solely because they are considered essential to the objective of this report.

ACKNOWLEDGMENTS

I extend special thanks to the members of R&D and our contractors for their extensive input and feedback through the entire process of developing this Evaluation Implementation Plan.

John Tunna
Director
Office of R & D
Federal Railroad Administration
U.S. Department of Transportation

CONTENTS

EXECUTIVE SUMMARY 1

1. Introduction 1

1.1 R&D Evaluation Mandate 3

1.2 R&D Evaluation Goals 3

1.3 R&D Evaluation Standards 3

2. Uses of Evaluation 5

2.1 Formative Evaluation 5

2.2 Summative Evaluation 5

3. Types of Evaluation 6

3.1 Context Evaluations 6

3.2 Input Evaluations 6

3.3 Implementation Evaluations 6

3.4 Impact Evaluations 6

4. Evaluation Framework 8

5. Key Evaluation Questions 9

6. Start-up Pilot Evaluations 10

6.1 Pilot Evaluation Execution and Funding 10

6.2 Building Evaluation Capacity and Credibility with Internal and External Evaluations 10

6.3 Program Evaluation Plans – Designing the Evaluations 15

6.4 Evaluating Pilots and periodic updating of R&D's evaluation practices 16

7. Resources Available to Support Evaluations 17

7.1 Evaluation Coordinator 17

7.2 Evaluation Contractors 17

7.3 Administrative and Clerical Support 18

7.4 Evaluation Budget 18

7.5 Evaluation Report Format 18

7.6 Evaluation Standards Attestation Form 20

8. Institutionalizing and Mainstreaming Evaluation 21

8.1 Metaevaluation 21

8.2 The Evaluation Manual 21

A. Appendix 22

A.1 Program Evaluation Definition 22

A.2 Guiding Principles of Sound Evaluation 23

A.3 Criteria for Assessing Different Stages of R&D 25

A.4 R&D Example Program Logic Models 26

A.5 Application of Evaluation Framework – Grade Crossing Research 28

A.6 Outline for the Evaluation Manual 29

A.7. Evaluation Standards Attestation Form 32

A.8 Evaluation Standards 35

a. *Utility standards* 35

b. *Feasibility standards* 35

c. *Propriety standards.* 36

d. *Accuracy standards* 36

e. *Evaluation accountability standards*.. 36

EXECUTIVE SUMMARY

This evaluation implementation plan was developed as a foundation for guiding systematic, improvement-oriented evaluations and institutionalizing program evaluation throughout the Office of R&D. The projected evaluations' key purposes are to:

- contribute to improving railroad safety,
- guide and strengthen the Office's programs,
- facilitate knowledge diffusion and technology transfer,
- meet R&D's accountability requirements, and
- build R&D's evaluation capacity.

The plan for fulfilling these objectives provides the Office of R&D's Director, Division Chiefs, staff, and contractors with a consistent, standards-based, practical approach to evaluating R&D projects and programs.

The approach is keyed to four types of evaluation. Evaluations of context, inputs, implementation, and impacts respectively ask: What needs to be done? How can it best be done? Is it being done? Did it succeed?

The approach is also keyed to two main uses of evaluation findings. *Formative* uses include systematic feedback for strengthening program activities, also for proactively identifying and preventing safety problems. *Summative* uses involve retrospective assessments and reporting of program outcomes—intended and unintended—to support data-driven decisionmaking, formulations of evidence-based policies, and public accountability.

This plan calls for a pilot stage to test, refine, and mainstream the new evaluation approach. Each R&D division has selected one or more current projects for pilot evaluation. Division Chiefs and program managers will be engaged to spell out evaluation questions that they see as especially important. Along the way they will be involved in ensuring that the evaluation is targeting and producing useful feedback for both program improvement and accountability.

As determined by the divisions and in consultation with R&D's Director, the pilot evaluations may be:

- internal evaluations, directed and funded by program and project management staff actively engaged in the development and implementation of the project,
- internal evaluations directed and funded by program and project management staff who are not actively engaged in developing or implementing the project, or
- external evaluations (called metaevaluations) of the internal evaluations.

The pilot evaluations will be coordinated and assisted by an evaluation coordinator to be designated by R&D's Director. In addition to coordinating the pilot evaluations, the evaluation coordinator will be available to provide each division with guidance and technical evaluation support should they need and request it.

Throughout the pilot stage, the evaluation coordinator will engage R&D's Director, Division Chiefs, program managers, and other senior staff to assess and help strengthen the involved evaluation approach. Key standards of the evaluation process will be used to guide the assessment and improvement of the individual pilot evaluations and the overall evaluation approach. Those standards include utility, feasibility, propriety, accuracy, and evaluation accountability.

Ultimately, the pilot evaluations will be evaluated against the standards, with the metaevaluation results being used to improve and help institutionalize the R&D Office's approach to program evaluations. The culmination of this process will be an evaluation manual for use throughout the Office of R&D.

1. INTRODUCTION

1.1 R&D Evaluation Mandate

The Government Performance and Results Act (GPRA) and the GPRA Modernization Act of 2010 require Federal agencies to assess the manner and extent to which their programs achieve intended objectives. Consistent with the GPRA directive, this plan was developed for increasing and improving program evaluation throughout the Office of R&D. In general, the intent is that evaluation should become integral to all R&D programs and, to at least a minimally acceptable degree, evaluation methods should be built into each project from the start.

1.2 R&D Evaluation Goals

R&D evaluation goals focus on five broad areas:

- Inform and assist *continuous improvement of program effectiveness and impact on railroad safety;*
- Contribute to long-term improvements in *knowledge diffusion and technology transfer;*
- Strengthen the Office of R&D's effectiveness, stature, credibility, and case for continued financial support;
- Meet *accountability* requirements; and
- Build *evaluation capacity* by making evaluation design, budgeting, and staffing a regular part of each division's R&D budget requests, program planning, procurements in contracts and grants, program implementation, and program accountability.

This plan calls for conducting and learning from pilot evaluations that:

- in combination, span the entire R&D life cycle, including program planning, basic research, development, implementation, diffusion, and technology transfer;
- are of interest and use to R&D's full range of intended stakeholders, including the Office of Railroad Safety, Congress, the railroad industry, and the public;
- employ relevant and applicable program performance metrics;
- effectively report findings to agency policy makers, senior safety leaders in the rail industry, and other key industry decisionmakers; and
- make credible, positive contributions to program decisionmaking and policy development.

1.3 R&D Evaluation Standards

The R&D Office's pilot evaluations—and ultimately all of its program evaluations—should be guided by standards of the evaluation profession. This plan recommends initial adoption of the standards contained in Joint Committee on Standards for Educational Evaluation (2011) *The Program Evaluation Standards* (Sage) for use as guiding principles when conducting evaluations. All pilot evaluations should apply the applicable standards and, at an evaluation's end, complete and append to the final report the

Evaluation Standards Attestation form which is included in Appendix A.7. See Appendix 8 – Evaluation Standards.

2. USES OF EVALUATION

Evaluation comprises two main uses, formative and summative:

2.1 Formative Evaluation

Formative evaluations are conducted before and during a program to guide planning and implementation and help ensure success. These evaluations are especially useful in helping those who plan and carry out programs to systematically identify and address emerging needs and issues as they arise. Formative evaluations primarily seek *to improve* a program.

2.2 Summative Evaluation

Summative evaluations are based on information collected throughout the course of an evaluation but are consummated at the end of a program or program cycle to take stock of accomplishments, understand a program's impact, and meet accountability requirements. Summative evaluations primarily seek *to prove* or demonstrate a program's success. Summative evaluations are especially useful to funding organizations, oversight bodies, client groups, and the public and should be keyed to their questions as well as fundamental values, such as railroad safety, risk reduction, and reduction in injuries and fatalities. Prior R&D summative evaluations of safety culture interventions, for example, were highly influential in the development of FRA's Risk Reduction Program and that part of the Rail Safety Improvement Act of 2008, influencing broad safety culture change in the industry overall.

3. TYPES OF EVALUATION

This plan is centered on four types of evaluations: context, input, implementation, and impact evaluation.

3.1 Context Evaluations

Context evaluations assess needs, problems, assets, and opportunities, plus relevant contextual conditions and dynamics. Decisionmakers use context evaluations to define goals and set priorities and to make sure program goals are targeted to address significant, assessed needs and problems. Oversight bodies and program stakeholders use context evaluation findings to judge whether the program was guided by appropriate goals and also to judge outcomes for their responsiveness to the program's targeted needs, problems, and goals.

3.2 Input Evaluations

Input evaluations assist program planning by identifying and assessing alternative approaches, competing action plans, staffing plans, and budgets for their feasibility and potential cost-effectiveness to meet targeted needs and achieve goals. Input evaluations also assist program planning with the design of evaluation and implementation plans. Decisionmakers use input evaluations to identify and choose among competing plans, identify possible program performance measures, write funding proposals, allocate resources, assign staff, schedule work, and ultimately help others judge an effort's plans and budget.

3.3 Implementation Evaluations

Implementation evaluations monitor, document, assess, and report on the implementation of program plans. Such evaluations provide feedback throughout a program's implementation and later report on the extent to which the program was carried out as intended and required. Program staff use periodic implementation evaluation reports to take stock of their progress, identify implementation issues, and adjust their plans and performance to assure program quality and on-time delivery of services. At the end of the program or after a program cycle, the program's staff, overseers, and constituents may use the implementation evaluation's documentation to judge how well the program was carried out. They may also use the implementation's documentation to judge whether a program's possibly deficient outcomes were due to a weak intervention strategy or to inadequate implementation of the strategy.

3.4 Impact Evaluations

Impact evaluations identify and assess costs and outcomes of a program or project—intended and unintended, short term and long term. These evaluations provide feedback during a program's implementation on the extent to which program goals are being addressed and achieved; at the program's end, impact evaluations identify and assess the program's full range of accomplishments. Program staff use interim impact evaluation feedback to maintain focus on achieving important outcomes and to identify and address deficiencies in the program's progress toward achieving successful outcomes. Ultimately,

impact evaluations assess and report on a program's accomplishments. Program overseers, funders, and constituents use final impact evaluation results to judge whether the program's accomplishments were significant and worth the cost. The key questions are: Did the program achieve its goals? Did it successfully address the targeted needs and problems? Were the outcomes worth the cost?

In summing up long-term evaluations, the impact evaluation component may be divided into four subparts of assessment: reach to the targeted communities or group of beneficiaries, effectiveness, sustainability, and transferability. These impact evaluation subparts ask: Were the right beneficiary groups reached? Were the targeted needs and problems addressed effectively? Were the gains in a program's accomplishments and mechanisms to produce them sustained and affordable over the long term? Did the strategies and procedures that produced the accomplishments prove to be transferrable, adaptable, and affordable for effective use elsewhere?

4. EVALUATION FRAMEWORK

Table 1 summarizes employment of the four types of evaluation for both formative and summative purposes. The matrix's eight cells encompass the main types of evaluative information needed to guide evaluation activities and produce credible, and therefore defensible, formative and summative evaluation reports. This table is intended to help evaluators in the Office of R&D conceptualize, plan, and conduct evaluations that serve both formative and summative roles and address a full range of important evaluative questions throughout the R&D lifecycle.

Table 1. The Relevance of Four Evaluation Types to Formative and Summative Evaluation Roles

Evaluation Roles	Types of Evaluation			
	Context	**Input**	**Implementation**	**Impact**
Formative: Proactive application of descriptive and judgmental information to assist decisionmaking, program implementation, quality assurance, and accountability	Guidance for identifying needed interventions, choosing goals, and setting priorities by assessing and reporting on needs, problems, risks, assets, and opportunities	Guidance for choosing a program strategy (and possibly an outside contractor) and settling on a sound implementation plan and budget by assessing and reporting on alternative strategies and resource allocation plans and subsequently closely examining and judging the operational plan and budget	Guidance for executing the operational plan by monitoring, documenting, judging, and repeatedly reporting on program activities and expenditures	Guidance for continuing, modifying, certifying, or terminating the program by identifying, assessing, and reporting on intermediate and longer term outcomes, including side effects
Summative: Retroactive use of descriptive and judgmental information to sum up the program's value, e.g., its quality, efficiency, cost, practicality, safety, impact, and significance	Judging goals and priorities by comparing them with assessed needs, problems, risks, assets, and opportunities	Judging the implementation plan and budget by comparing them with targeted needs, problems, and risks; contrasting the plan and budget with critical competitors; and assessing their compatibility with the implementation environment and compliance with relevant codes, regulations, and laws	Judging program execution by fully describing and assessing the actual process and costs, comparing the planned and actual processes and costs, and assessing compliance with relevant codes, regulations, and laws	Judging the program's success by comparing its outcomes and side effects to targeted goals, needs, problems, and risks; examining its cost-effectiveness; and, as feasible, contrasting its costs and outcomes with competitive programs; also interpreting results against the effort's outlay of resources and the extent to which the operational plan was both sound and effectively executed

5. KEY EVALUATION QUESTIONS

Table 2 summarizes the types of formative and summative questions to be addressed by context input, process, and product evaluations. This matrix poses example evaluation questions focused on improving safety for the railroad industry.

Table 2. Types of Formative and Summative Questions to Be Addressed by Context, Input, Process, & Product Evaluations

Evaluation Roles	Types of Evaluation			
	Context	**Input**	**Implementation**	**Impact**
Formative	- What are the highest priority research and evaluation needs in given program areas of the Office of R&D?	- Given the assessed priorities, what are the most promising potential programs for producing the needed research and evaluation? - How do these alternatives compare (potential success, costs, etc.)? - How can the needed research and evaluation be most effectively designed, staffed, funded, and implemented? - What might be some barriers to effective implementation?	- To what extent is the R&D program proceeding on time, within budget, and effectively? - If necessary, how can the design be improved? - How can one strengthen the implementation?	- To what extent are indicators of success being observed and assessed? - What other indicators, if any, have emerged that show the program is or is not succeeding? - What side effects (positive or negative) are emerging? - How can the implementation be modified to maintain success?
Summative	- To what extent did R&D in this program area address high priority needs? - To what extent did program goals reflect the targeted assessed needs?	- What research strategy was chosen and why, compared with other viable strategies (e.g., prospects for success, feasibility, costs)? - How well was the chosen strategy converted to a sound, feasible work plan?	- To what extent was the program carried out as planned or modified with an improved plan? - How well was the program executed?	- To what extent were the originally assessed R&D program goals and needs effectively addressed/achieved? - Were there any unanticipated negative or positive side effects? - What conclusions can be reached (i.e., cost effectiveness, sustainability, applicability)?

6. START-UP PILOT EVALUATIONS

Key evaluation questions will be developed and refined for each of the division's selected in-depth pilot evaluations. The pilot evaluations will be conducted and utilized as a key means to testing, refining, and institutionalizing a sound, feasible evaluation approach throughout the Office of R&D. These in-depth evaluations will also provide R&D staff members with real world-based evaluation capacity development experiences. The four division's selected pilot evaluations are highlighted in Table 3.

Table 3. In-Depth Pilot Evaluations by Division

Division	Program(s) to be Evaluated
Rolling Stock	• Rail Energy, Environment, and Engine-Efficiency Technology Research Program
Track	• Neutral Temperature and Incipient Buckling Detection System for Continuous Welded Rail • Autonomous Track Geometry Measurement System (ATGMS) Evaluation
Train Control & Communication	• Florida Trespass Prevention Research Program Evaluation
Human Factors	• Safety Culture Leadership Development • Safety Culture Impact Evaluation • Joint Human Factors and Train Control & Communication Research Program Evaluation

6.1 Pilot Evaluation Coordination

The Director of R&D will appoint and assign an evaluation coordinator to help develop the pilot evaluation plans and coordinate their execution. Pilot evaluation plans will be developed during FY2014 with the intent to begin execution of the plans shortly thereafter. In addition to coordinating the pilot evaluations, the evaluation coordinator will provide the staffs of the pilot evaluations with technical support or training should staff need and request it.

6.2 Building Evaluation Capacity and Credibility with Internal and External Evaluations

Both internal and external evaluations will be conducted for the pilot evaluations, depending on need as determined by the Division Chiefs and in consultation with R&D's Director and evaluation coordinator. Integrating the use of highly credible internal and external evaluations is an important strategy for building evaluation capacity and credibility and developing evaluation-oriented researchers within FRA R&D.

Internal evaluations are carried out by evaluators internal to the organization or program being evaluated (i.e., FRA or Volpe evaluators). To minimize bias and to the extent possible, internal evaluators should be independent of the program or project they are evaluating. However, highly useful and credible evaluations can still be conducted when

evaluators are directly responsible for managing or implementing the program(s) being evaluated. This is especially true when the internal evaluator has considerable content and/or evaluation expertise, uses evaluation standards, and is perceived by key stakeholders as highly credible and trustworthy.

External evaluations are carried out by evaluators external to the organization or program being evaluated (i.e., evaluation contractors), ideally free of control or influence by those responsible for the design and implementation of the project or program. While often more costly and time consuming than internal evaluations, external evaluations are typically perceived by outside organizations as more objective and credible.

Table 4 below highlights some of the key differences, including advantages and disadvantages, between internal and external evaluations.

Table 4. Internal and External Evaluation

	Internal Evaluation	External Evaluation
Definition:	*An evaluation carried out by evaluators internal to the organization or program being evaluated.*	*An evaluation carried out by evaluators external to the organization or program being evaluated, ideally free of control or influence by those responsible for the design and implementation of the project or program.*
Evaluator characteristics:	- Internal staff of the organization, committed to organizational improvement - Ideally, independent of the program being evaluated - Sometimes, has commitment to the success or failure of the project or program being evaluated - Periodic or ongoing responsibility for evaluation in the organization - Usually has more content and organization expertise than external evaluators - Usually has more knowledge about organizational and industry context - Established internal working relationships - Ability to communicate timely and relevant evaluative information to key staff	- External staff to the organization funding the project or program, including consultants, contractors or grantees and their staff - Independent of the program and/or organization being evaluated - Usually has more evaluation-related expertise than internal evaluators - Less commitment to the success or failure of the project or program being evaluated - Less likely to be influenced by job or organizational factors
Advantages:	- Timeliness: Information often available immediately by informal briefings to inform program/policy decisions - Buy-in: Internal evaluators can rapidly involve key stakeholders and can facilitate implementation and evaluation use more quickly - Appropriateness: Evaluation methods more likely to be tailored to organization needs - Low cost: Cost is usually less than external - Sustainability: Internal evaluators often considered an enduring	- Objective: External evaluations often perceived as more objective than internal evaluations - Credible: External evaluations often perceived as more credible than internal evaluations - Expertise: Specialized evaluation skills and evaluation teams often more readily available with consultants. Evaluation consultants also may have more exposure to a wider range of evaluations, methods and practices than internal evaluators. - More formal reports/briefings - Provides learning opportunities for

	organization resource	internal evaluators
Disadvantages:	- Time consuming: Staff time to conduct internal evaluations may be limited - Bias: Internal evaluations can be perceived as more biased than external evaluations - Lower expertise: Often impractical to hire staff with highly qualified evaluation skills	- Higher cost - Longer timeframe for reports/briefings - Oversight and control more difficult or challenging - Technical expertise may be more difficult to find in specialized content areas
Primary Intended Users:	- Program staff - Senior managers - Key organization decisionmakers	- Funders - Legislators - External organizations/individuals
Primary Intended Uses:	*Primarily formative* - Improvement-oriented - Organizational efficiency and effectiveness - Program planning, performance monitoring, and strategy development - Management, policy, and key decisionmaking support Secondarily summative - Documentation for accountability - Senior-level decisionmaking - Reporting to FRA, DOT, OMB, and Congress	*Primarily summative* - Accountability - Reporting to FRA, DOT, etc. - *Secondarily formative* - Rethinking programs - Setting Office of R&D priorities - Strategic planning
Examples:	Evaluation of a research product, tool or intervention by internal staff	-Evaluation contractors' evaluations -Inspector Generals' audits -GAO evaluations -TRB Reviews of R&D Program

The following guidelines should be used to decide whether pilot evaluations of a contracted project will be an internal or external evaluation:

a) an exclusively internal, self-evaluation conducted by the funded project contractor or grantee staff responsible for designing and implementing the project,

(acceptable when the project contractor or grantee has a track record of conducting sound, useful evaluations, and when the involved R&D division chief is confident that the funded project will credibly evaluate its own work)

b) an internal self-evaluation conducted by the funded project contractor or grantee staff responsible for designing and implementing the project, with the applicable R&D division also contracting for an independent metaevaluation of the evaluation

(desirable when the applicable R&D division chief is confident that the contractor or grantee staff of the funded project will credibly evaluate its own work, but—under a condition of "trust but verify"—also wants an independent metaevaluation to validate the evaluation's quality)

c) an R&D Office-contracted external evaluation of a funded project

(acceptable and desirable, especially when the applicable division chief believes that key stakeholders stress the need for an independent evaluation conducted by a contractor or grantee external to the project being funded)

d) an R&D Office-contracted external evaluation of a funded project with the applicable R&D division also contracting for an independent metaevaluation of the evaluation

(desirable when the involved R&D division chief and key stakeholders deem the appearance and fact of evaluation independence to be essential for acceptance and use of the evaluation report)

e) an internal evaluation of an R&D Office-funded project conducted by an evaluator in the R&D Office

(desirable when a qualified evaluator in the R&D Office is available to conduct the evaluation, when internal evaluation support can be provided, when proceeding accordingly would help build evaluation capacity within the R&D Office, or when a timely evaluation is important)

f) an internal evaluation of an R&D Office-funded project by an evaluator in the R&D Office, with the R&D Office also contracting for an independent metaevaluation of the evaluation

(desirable when a qualified evaluator in the R&D Office is available to conduct the evaluation or when internal evaluation support can be provided, when proceeding

accordingly would help build evaluation capacity within the R&D Office, and when key stakeholders expect the evaluation findings to undergo independent verification)

Depending on the judgments of R&D's Director and Division Chiefs, any of the above arrangements for an evaluation may be acceptable or even desirable. It is crucial that the division chiefs—in their role of evaluation-oriented researchers—lead in deciding how best to arrange for evaluations.

6.3 Program Evaluation Plans – Designing the Evaluations

The following outline is to assist R&D's evaluators in developing sound designs for evaluating R&D programs and should be used as a general template when developing program evaluation plans:

- Introduction
  - Description of the program to be evaluated
  - The evaluation's client and other key evaluation audiences
  - Intended uses of evaluation findings (formative and/or summative)
- The Evaluation's Advance Organizers
  - Type(s) of evaluation needed (context, input, implementation, and/or impact)
  - Key evaluation questions (keyed to context, input, implementation, and/or impact evaluations)
- Data Collection Plan
  - Needed information (quantitative and qualitative; existing and new)
  - Sources of information
  - Data collection tools
  - Data collectors
  - Timeline
- Organization and Analysis of Information
  - Proofing, coding, storage, and retrieval
  - Quantitative analysis
  - Qualitative analysis
- Reporting
  - Audiences
  - Interim reports
  - Final report
  - Due dates
- Metaevaluation
  - Standards for guiding and judging the evaluation (utility, feasibility, propriety, accuracy, and evaluation accountability)
  - Internal metaevaluation (documenting, assessing, controlling, and attesting to the evaluation's quality)
  - External metaevaluation (independent validation of the evaluation's findings)
  - Stakeholder engagement (examples)
  - Stakeholder review panel (SRP)
  - Webinars
  - Feedback sessions

Focus groups
Public forums

Management of the Evaluation

Staff
Consultants
Schedule (e.g., a Gantt chart)
Evaluation budget (personnel, travel, consultants, supplies, communications, services, indirect costs, subcontracts)

6.4 Evaluating Pilots and Periodic Updating of R&D's Evaluation Practices

Once the pilot evaluations have begun, they should be studied to shed light on their implications for R&D's evaluation practices in general. Questions to be addressed through such study may include the following:

- How applicable and useful were the program evaluation standards for guiding and judging the pilot evaluations?
- How could the standards be adapted for sustained use throughout the Office of R&D?
- To what extent did the pilot evaluations contribute to improving R&D railroad safety?
- How useful and in what ways were the pilot evaluations useful for formative purposes?
- How useful and in what ways were the pilot evaluations useful for summative purposes?
- To what extent and with what levels of utility did the pilot evaluations apply all four types of evaluation: context, input, implementation, and impact?
- To what extent did the pilot evaluations make use of and find useful independent audits or metaevaluations?
- To what extent did the pilot evaluations prove to be cost-effective?
- What were the most important lessons from the pilot evaluations?

7. RESOURCES AVAILABLE TO SUPPORT EVALUATIONS

7.1 Evaluation Coordinator

This plan provides for engagement of an evaluation coordinator to facilitate the pilot evaluations. The coordinator's primary responsibility is to facilitate the pilot evaluations and provide their staff members with technical assistance and training, as needed. Under this assignment, the evaluation coordinator will serve the role of Contracting Officer's Representative while a program manager serves as Task Monitor.

The coordinator will facilitate collaboration across different evaluation studies and program areas to effect efficiencies in data collection and effectiveness in communicating aggregate findings. The evaluation coordinator will also regularly interact with the Division Chiefs and the Director of R&D to identify and help address common information needs (e.g., by coordinating data collection and common issues across different evaluations). In addition, the evaluation coordinator will exploit opportunities to combine evaluation findings from different evaluations in generalized reports for issuance by the Office of R&D. The coordinator will attend and participate in weekly meetings of Division Chiefs, as appropriate, to facilitate coordination of pilot evaluations across divisions.

7.2 Evaluation Contractors

Where needed, this plan provides for external contractors to conduct or assist the pilot evaluations. The evaluation coordinator and Division Chiefs, in coordination with R&D's Director, will deliberate and decide on whether, when, and how to engage such external evaluation resource personnel, depending on the divisions' needs for augmenting their evaluation capabilities. Such external evaluation services may take the form of, for example, constructing evaluation instruments, analyzing data, issuing independent reports, conducting and reporting a metaevaluation or peer review, assisting in solving particular problems that arose in an evaluation, or conducting the evaluation.

Statements of work for evaluation tasks should include:

- Descriptions of evaluation purposes and questions
- Right-to-know audiences and their intended uses of findings
- Standards for guiding and judging the evaluation
- Staffing of the evaluation
- Required information
- Protocols for gathering and handling the information
- Required reports and editorial authority
- Dissemination of reports
- Documentation of the evaluation procedures and information
- Provisions for metaevaluation or peer review
- Budget for the work
- Schedule and conditions for payment

7.3 Administrative and Clerical Support

In order for the evaluation implementation plan to be effectively implemented, necessary clerical and administrative support staff will need to be in place. In addition to the evaluation coordinator, clerical and administrative staff support is currently available through existing contractors. Careful consideration should be given to choosing evaluation support staff, as they can serve an important role in helping build evaluation capacity internally in FRA and/or externally with contractors, including Volpe.

7.4 Evaluation Budget

The projected in-depth start-up evaluation pilots in each division provide the foundation for R&D's evaluation capacity development effort. This plan calls for funding at least one in-depth pilot evaluation in each division (approximately $300,000 per evaluation over a 3-year period). It is also recommended that the Director of R&D and the Division Chiefs consider the desirability and feasibility of adding a pilot evaluation in the area of High Speed Rail. It is assumed that the Office of R&D will also allocate sufficient resources for such support areas as coordination, clerical/administrative support, and technical evaluation support.

7.5 Evaluation Report Format

The following is a model outline for R&D and Volpe staff members to consider in writing final program evaluation reports.

Outline for a Final Program Evaluation Report

Prologue (Origin of the program evaluation)
- Who requested the evaluation, why, and for whom?
- Who are the evaluators, what are their perspectives and credentials, how did they approach the assignment?
- What is the subject program's title?
- What is the program's mission?
- What are this evaluation's bottom-line questions?

Introduction (National significance of the subject program and overview of the report)
- What national needs and problems provide the focus for the program?
- How has this evaluation documented the program's approach and impacts?
- What are the key audiences for the evaluation?
- How is the report organized to address the differential needs of these audiences?

Background of the Program (Descriptive and intended for use by all audiences)
- What group founded program, when, and why?
- What are the program's goals?
- Who are the intended beneficiaries of program services?
- What is the administrative structure of the program?
- What is the social and political context in which the program operates?
- Photographic reprise to depict key aspects of the program's background

Program Implementation (A strictly descriptive account, intended especially for those who might be interested in replicating the program's approach)

- Overview of the program
- Management and coordination
- Development of program protocols and procedures
- Collaborative arrangements
- Staff assignments
- Metrics and data collection
- Funding
- Internal and external communication
- Review and revision
- Photographic reprise to depict key aspects of the program's operations

Results (Evaluative and intended especially for oversight bodies and a wide range of interested audiences)

- Approach to assessing the program's quality, importance, and cost-effectiveness
- Context Evaluation: Are the program's goals addressed to the railroad industry's important needs, problems, and opportunities?
- Input Evaluation: Is the program's approach maximally responsive to assessed and targeted needs and problems in the railroad industry?
- Implementation Evaluation: Has the program's administrators and staff effectively implemented the program's plan of action?
- Impact Evaluation: What is the extent and significance of the program's positive outcomes, negative outcomes, and unintended side effects?
- Sustainability Evaluation: To what extent are the program's successful practices and positive outcomes being sustained?
- Transferability Evaluation: To what extent has the program's approach been successfully adapted and applied elsewhere?
- Photographic reprise to highlight and make vivid the program's accomplishments

Conclusions (Intended for all audiences)

- The program's notable strengths
- The program's notable weaknesses
- Key lessons learned
- Bottom-line assessment of the program's merit and worth
- Photographic reprise depicting the evaluation's main message

Appendix

- References
- Key data sources and tools
- About the evaluators
- Attestation of the evaluation's adherence to standards of utility, feasibility, propriety, accuracy, and accountability (employing the form in Appendix A.7)
- Members of the Evaluation Review Panel

7.6 Evaluation Standards Attestation Form

The bottom line requirement for the pilot program evaluations is that they meet professionally defined standards of sound evaluation. To support that purpose, an Evaluation Standards Attestation Form has been developed and is included in Appendix A.7. In general, the form is keyed to the five main categories of standards in the ANSI-approved Joint Committee (2011) *Program Evaluation Standards*. Those categories are Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability, which are defined in Appendix A.8. The form contains summaries of the 30 specific standards found in the 2011 *Program Evaluation Standards.*

Evaluators of the pilot programs are expected to plan and conduct their evaluations in accordance with the 30 standards, although some of the standards may not be applicable to particular evaluations. At the evaluation's conclusion, the evaluator is expected to complete a copy of the Evaluation Standards Attestation Form and append it to the final report. In completing the form the evaluator should judge and assess whether each standard was met, partially met, or not met, or whether the standard was judged not to be applicable. In addition to marking the form accordingly, the evaluator should provide one or two sentences of justification for the marked judgment. By gauging the evaluation to meet the 30 standards from the start, the evaluator will enhance prospects for the evaluation's success. By completing the Evaluation Standards Attestation Form at the evaluation's end and appending the form to the final evaluation report, the evaluator will be demonstrating his or her professionalism and accountability for completing a creditable, useful evaluation.

8. INSTITUTIONALIZING AND MAINSTREAMING EVALUATION

The ultimate aim of this evaluation capacity development effort is to make systematic evaluation an integral part of all programs in the Office of R&D, such that evaluation contributes to program success, meets the Office's accountability needs, and provides direction for continual improvement of program planning, execution, and influence in strengthening safety throughout the railroad industry. This evaluation capacity development effort will culminate in two main activities: metaevaluation and the development and adoption of an evaluation manual.

8.1 Metaevaluation

Metaevaluation is the evaluation of evaluation based on professionally developed standards of sound evaluation practices. With the completion of each pilot evaluation, a detailed metaevaluation checklist keyed to the 2011 Joint Committee, ANSI-approved *Program Evaluation Standards* will be applied to evaluate the subject evaluation's strengths and weaknesses. The metaevaluation results from all of the pilot evaluations will then be used to strengthen and finalize the Office's evaluation approach.

8.2 The Evaluation Manual

The final version of the Office's evaluation approach will be documented in an evaluation manual for use throughout the Office of R&D. This evaluation manual will be the main tool for facilitating systematic evaluation in R&D's full range of programs. Periodically, the manual will itself undergo evaluation and revision, as needed. A working outline of the evaluation manual appears in Appendix A.6.

APPENDIX

A.1 Program Evaluation Definition

At its most basic level, program evaluation is an assessment, both formatively and summatively, of a program's value. Specifically, program evaluation is the systematic process of delineating, obtaining, reporting, and applying descriptive and judgmental information about such aspects of a program as its relevance, quality, efficiency, safety, legality, effectiveness, impact, sustainability, transportability, cost-benefits, significance, etc. Evaluation may be conducted internally, as in self-evaluations, or by external evaluators. It serves those who fund, oversee, and conduct a program and those who use its products and services. A comprehensive program evaluation addresses questions about needs and problems, program goals and priorities, program strategies and designs, program implementation, program costs, and program results and impacts.

Program evaluations produce both interim and final reports, formal and informal, employ both qualitative and quantitative methods, and incorporate a variety of methodological approaches, including but not limited to:

- needs assessments
- stakeholder engagement strategies (review teams, feedback sessions)
- case studies
- structured and semistructured interviews
- focus groups
- surveys
- cost benefits analysis
- logic modeling
- field observations
- quasi experiments
- randomized, controlled, comparative studies

At its best, a program evaluation meets professional standards for sound evaluations. Main uses of program evaluations are to help focus, guide, and strengthen programs; issue accountability reports; help the adoption of effective practices and technology; provide a basis for funding requests; and, as appropriate, make decisionmakers, stakeholders, and consumers aware of programs that proved unworthy of further use. Program evaluation is a ubiquitous process that applies across disciplines and national boundaries. Employing and using improvement-oriented evaluation is the essence of professional service and responsibility. Moreover, sound evaluation is one of society's most powerful forces for ensuring program success and holding program providers accountable for efficient and effective use of resources to serve the public good.

A.2 Guiding Principles of Sound Evaluation

The following guiding principles undergird all aspects of the evaluation implementation plan:

- **Utilization-Focused Evaluations**. Systematic evaluations should be funded and undertaken only if there is a clear understanding of the intended users of the evaluation and its intended uses—and when it is clear that the intended users need and will apply the findings in intended ways.

  This plan strongly recommends that the R&D Office's evaluation practices not become bureaucratic rituals in which costly reports on the full range of projects are produced, rarely used, and only sit on shelves gathering dust. Instead, priorities for funding and conducting evaluations should be assigned to those evaluations that have a clear stakeholder engagement and buy-in and a strong prospect for helping to improve the conduct and use of R&D programs.

- **Impact and Improvement Oriented Evaluation.** The most important purposes of evaluation are not only to prove but to improve.

  By conducting state-of-the-art evaluations for all of its programs, R&D will not only document program impact but also constantly strengthen its efforts to improve the safe, efficient, and reliable movement of people and goods in the U.S. railroad industry.

- **Priority Evaluations**. To make best use of available resources for evaluations, evaluation services will be focused on the highest priorities for evaluative feedback.

  In general, evaluation should be integral to all R&D programs and, to at least a minimally acceptable degree, evaluation methods should be built into each project from the start. However, resources for evaluation are limited and some needs for evaluation are more important than others. Consequently, the R&D Office will apply criteria and guidelines to set priorities and assign time and funds in accordance with the judged importance of proposals and requests for evaluation. This stipulation applies particularly to evaluations demanding a high degree of sustained, in-depth study.

- **Cost-Free Evaluations.** In general, evaluations should, in the long run, be cost-free by virtue of making programming more efficient (thus saving time and money) and more effective and accountable (thereby strengthening R&D's case for sustained and increased funding).

  It is patently clear that taxpayers and their representatives in Congress expect government-funded programs to use money wisely, properly, frugally, and effectively and demonstrate their fiscal accountability and positive service to society through systematic, unbiased evaluations. Employment of sound, systematic evaluation processes is essential to foster success, help keep allocations and expenditures at appropriate levels, and meet government and public requirements for accountability. It is posited that R&D's returns on investment in systematic evaluation will over time produce results and attract funding whose value equals or exceeds the costs of the evaluations. This is suggested as a worthy goal, and R&D is

encouraged to regularly test this hypothesis by including cost-effectiveness of their evaluations as a key criterion for assessing and over time strengthening its evaluations.

A.3 Criteria for Assessing Different Stages of R&D

Because R&D's contributions span the full range of processes involved in improving railroad operations, their evaluations will address a wide range of success criteria, such as the following:

- **Basic Research**

  Reliable information, rigorous process, internal validity, and external validity

- **Development**

  Face validity/appropriateness, estimated viability, projected impact and relative contribution, cost, feasibility, tractability, operability, and efficiency

- **Translation, pilot, and dissemination**

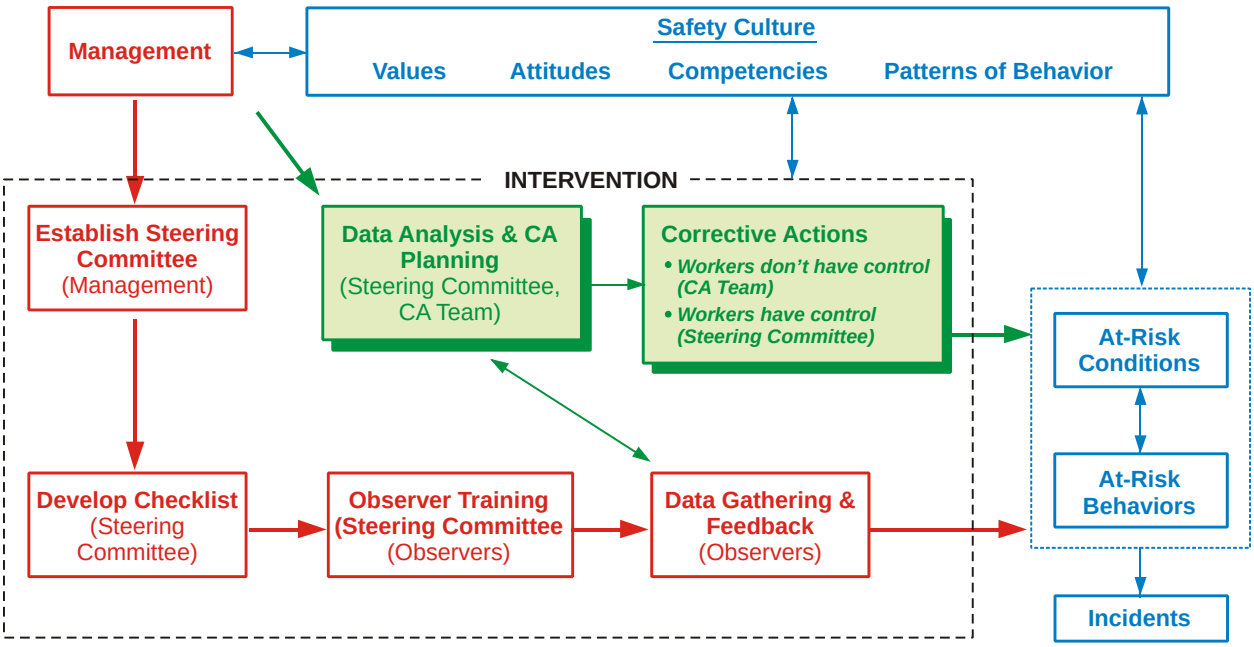
  Clarity and truthfulness of message; feasibility, practicality, relevance of product or process; pervasiveness in communicating with potential users; demonstrated impacts on key targets

- **Technology Transfer and Stakeholder Adoption**

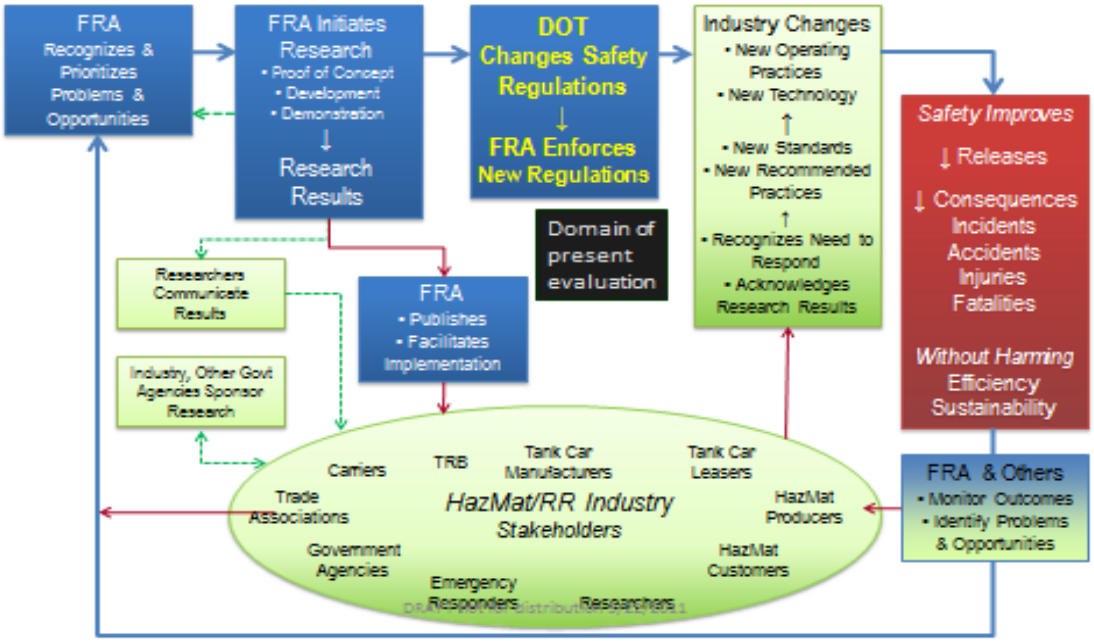
  Quantity, continuity, capabilities, motivation, and proficiency of trained users of the product or process; adaptability and ease of use of the product or process; adoption and use by intended users; cost-effectiveness of field applications of the product or process; sustained use of the product or process, spread of the product's or process's use; valuation and support of the product or process by the targeted users; integration of the product or process into the users' programs; meaningful feedback from users aimed at future improvements of the employed products and processes

A.4 R&D Example Program Logic Models

Clear Signal for Action (CSA) Theory of Change



HAZMAT Research Program Theory of Change



A.5 Application of Evaluation Framework – Grade Crossing Research

USDOT FRA R&D: Human Factors Division EVALUATION PLAN for Program to Improve Safety at Grade Crossings (discussion draft by DLS 8/8/12)				
Goals: Fewer deaths & accidents; Reduced Trauma; Cost/effective interventions; Power Structure Buy-in				
Stakeholder Engagement: Target Partners & Users	**Context Evaluation (What needs to be done?)**	**Input Evaluation (How should it be done?)**	**Implementation Evaluation (Is it being done?)**	**Impact Evaluation (Did it succeed?)**
• Establish a Review Panel: Researchers (TTech, Volpe, Weststat); Operation Lifesaver, Inc.; Unions (UTU, LET, signalmen); Highway engineers; FRA/RRS • Arrange for the Review Panel to examine & give feedback on drafts of evaluation plans, instruments, & reports • Plan to keep the broader community of stakeholders informed as appropriate concerning the development, implementation, & outcomes of the Grade Crossings Safety Improvement Program • Plan & budget to secure maximum beneficial uses of the evaluation's findings & implementation of successful aspects of the program	• Determine assignments for conducting the context evaluation & acquire pertinent financial commitments • Obtain & analyze accident statistics (incidence, severity, environmental influences, & causal factors) • Study motorists' behavior • Study bicyclists' & pedestrians' behavior for additional clues • Study existing safety mechanisms, including standards, education, enforcement, & compliance • Study highway construction standards & implementation by state DOTs • Study state drivers manuals & enforcement • Review data on track inspection practices & results • Identify relevant, key change agents • Identify constraints & barriers to corrective & improvement measures • Identify assets & opportunities for change • Create, test, & refine a user-friendly model of driver behavior • Provide the Review Panel and other pertinent stakeholder groups with the draft Context Evaluation report about 10 days in advance of a face to face feedback session • Conduct the feedback session, both to provide advance feedback of findings & to identify possible areas of inaccuracy or ambiguity • Use obtained feedback to finalize Context Evaluation report on needs, problems, opportunities, & environ factors related to improving grade crossings safety • Engage & inform key decision makers of the context evaluation results • Propose intervention program, including measurable outcomes	• Determine assignments for conducting the input evaluation & acquire pertinent financial commitments • Establish 2 or 3 independent design teams • Develop required contents for competing intervention plans (e.g., design & engineering of train crossing modifications, needed changes to state drivers' manuals and enforcement, proposed highway construction standards, recommendations for state DOTs' implementation of standards, and modifications of train horns & crossings rumble strips • Develop criteria for judging competing plans (e.g., coverage of required content, response to context evaluation findings, feasibility, cost, benefits, potential for stakeholder buy-in, resourcefulness in using available resources, compatible with related efforts) • Provide orientation and background information to each design team, including the user-friendly model of driver behavior • Obtain completing plans • Evaluate competing plans against the pre-established criteria, including projected costs & benefits • Study possibility of converging best features of the competing plans • Draft the final plan & obtain feedback from the Review Panel • Finalize the plan, including cost projections • Secure approval & funding for implementation of the Grade Crossings Safety Improvement Plan, at least for field testing the plan	• Determine assignments for conducting the process evaluation & acquire pertinent financial commitments • Identify & make arrangements to provide formative feedback to key decision makers & other stakeholders on a periodic basis • Monitor & document the implementation of the Grade Crossings Safety Improvement Program • Periodically, provide the Review Panel and other pertinent stakeholder groups with draft process evaluation reports about 10 days in advance of a face to face feedback session • Conduct periodic feedback sessions with the Review Panel & the other pertinent stakeholder groups • Use the obtained feedback to finalize a documentary report on the extent and quality of the implementation and cost of the Grade Crossings Safety Improvement Program	• Determine assignments for conducting the product evaluation & acquire pertinent financial commitments • Determine the extent to which the originally assessed & targeted needs were met by the Grade Crossings Safety Improvement Program • Conduct a Goal-free evaluation to identify and report on the program's unexpected positive & negative outcomes • Conduct a Success Case Method study to identify and highlight noteworthy successes of the program • Compile a draft summative evaluation report to pull together the context, input, process,& product evaluation findings • Provide the Review Panel and other pertinent stakeholder groups with the draft summative evaluation report about 10 days in advance of a face to face feedback session • Conduct the feedback session, both to provide advance feedback of findings & to identify possible areas of inaccuracy or ambiguity • Use the obtained feedback to finalize the summative evaluation report • Prepare and deliver to stakeholders user-friendly executive summaries, feedback sessions, Web-based reports, etc. to get the word out on the program's results • Engage relevant decision makers to promote impacts of the program's successful & needed features • Propose follow-up evaluations of the program's impacts, sustainability, & transportability
Key the evaluations to meeting standards of utility, feasibility, propriety, accuracy, & evaluation accountability.				
Engage independent evaluator or team – with expertise in evaluation and background in rail industry – to conduct & report formative & summative metaevaluations (against standards of utility, feasibility, propriety, accuracy, & accountability). A metaevaluation is an evaluation of an evaluation.				

A.6 Outline for the Evaluation Manual

MANUAL FOR PROGRAM EVALUATION
Office of Research and Development
Federal Railroad Administration

OUTLINE
Draft: 8-27-2013

1. Purpose of the Manual

This manual provides personnel of the Office of R&D with a framework, standards, procedures, checklists, and exemplars for planning, conducting, reporting, and applying findings from sound evaluations of R&D's programs.

2. R&D's Rationale for Program Evaluation

As directed by Dr. John Tunna in 2013, evaluation is to be integral to all R&D programs and, to at least a minimal degree, evaluation methods should be built into each project from the start. His directive responded to the Government Performance and Results Act (GPRA) and the GPRA Modernization Act of 2010 requiring Federal agencies to assess the manner and extent to which their programs achieve intended objectives.

In general, the purpose of R&D's program evaluations is not only to prove but to improve its programs. Thus, evaluations must be designed first and foremost to help focus, plan, and conduct programs that make substantial contributions to improving railroad safety. In addition, R&D will employ its evaluations to meet public requirements for program accountability and to maintain a record of R&D's accomplishments, including its programs' strengths and weaknesses. R&D will use its repository of completed evaluations to inform future programs, so that they build on lessons learned from previous programs.

3. R&D Evaluation Goals

R&D's evaluation goals focus on five broad areas:

- Inform and assist continuous improvement of program effectiveness and impact
- Meet accountability requirements
- Contribute to long-term improvements in knowledge diffusion and technology transfer
- Build evaluation capacity by making evaluation design, budgeting, and staffing a regular part of each division's R&D budget requests, program planning, procurements in contracts and grants, program implementation, and program accountability
- Strengthen the Office of R&D's effectiveness, stature, credibility, and case for continued financial support

4. Standards to Assure the Soundness of Program Evaluations

Evaluators of R&D programs are expected to apply rigorous standards to guide and assess its program evaluations. The categories of standards are Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability. Appendix A provides summaries for each category of standards

Appendix B contains a form for evaluators to use in applying and reporting on their use of the standards. This form provides an evaluator a means to attest to an evaluation's adherence to each of 30 specific standards. Evaluators are advised to consider all 30 standards in all stages of an evaluation: focusing, design, budgeting, contracting, conducting, reporting, and using findings.

At the end of an evaluation, the evaluator should append a completed Evaluation Standards Attestation Form to their evaluation report. On the form, the evaluator should place X's denoting whether each standard was met, partially met, or not met. In addition, in the spaces provided the evaluator should include a brief statement of justification for the placement of each X.

5. Uses of Program Evaluations

Evaluation comprises two main uses: *formative* to help focus, plan, monitor, and assure a program's effectiveness, and *summative* to assess the program's quality, outcomes, and significance.

6. Stakeholder Engagement

7. Criteria for Assessing Different Stages of R&D

8. Evaluation Framework

9. Designing Evaluations

10. Key Evaluation Questions

11. Staffing Evaluations

12. Budgeting Evaluations

13. Evaluation Statements of Work

14. Logic Models

15. Data Collection

16. Analysis and Synthesis of Findings

17. Interim Reporting

18. Final Reports

19. Metaevaluation/Audits of Evaluations

APPENDIX

A. Summary of Evaluation Standards

B. Evaluation Standards Attestation Form

C. Stakeholder Engagement Checklist

D. Evaluation Design Checklist

E. Evaluation Budgeting Checklist

F. Data Collection Devices Checklist

G. Interim Reporting Checklist

H. Format for a Final Report

I. Metaevaluation Checklist

A.7. Evaluation Standards Attestation Form

Evaluation Standards Attestation Form
DRAFT 8-22-2013[1]

Evaluators of R&D's programs should complete a copy of this form and append it to their final report, as an attestation of the extent to which the evaluation adhered to applicable, specific standards of Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability.

The following summaries of ANSI-approved standards—drawn from Joint Committee on Standards for Educational Evaluation (2011). *The Program Evaluation Standards.* Los Angeles, CA: Sage.—are reprinted with the Committee's authorization and have been adopted for use by FRA's Office of R&D.[2]

Standard	Standard Statements	Basis for Judgment	Judgment			
			Met	Partially Met	Not met	N/A
U1 Evaluator Credibility	*Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.*					
U2 Attention to Stakeholders	*Evaluations should devote attention to the full range of individuals and groups invested in the program or affected by the evaluation.*					
U3 Negotiated Purposes	*Evaluation purposes should be identified and revisited based on the needs of stakeholders.*					
U4 Explicit Values	*Evaluations should clarify and specify the individual and cultural values underpinning the evaluation purposes, processes, and judgments.*					
U5 Relevant Information	*Evaluation information should serve the identified and emergent needs of intended users.*					
U6 Meaningful Processes and	*Evaluation activities, descriptions, findings, and judgments should encourage use.*					

[1] The designations U, F, P, A, and E, respectively refer to categories of standards labeled Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability, detailed in Appendix A.8.

[2] The designations U, F, P, A, and E, respectively refer to categories of standards labeled Utility, Feasibility, Propriety, Accuracy, and Evaluation Accountability, detailed in Appendix A.8.

Products						
U7 Timely and Appropriate Communicating and Reporting	*Evaluations should attend in a timely and ongoing way to the reporting and dissemination needs of stakeholders.*					
U8 Concern for Consequences and Influence	*Evaluations should promote responsible and adaptive use while guarding against unintended negative consequences and misuse.*					
F1 Project Management	*Evaluations should use effective project management strategies.*					
F2 Practical Procedures	*Evaluation procedures should be practical and responsive to the way the program operates.*					
F3 Contextual Viability	*Evaluations should recognize, monitor, and balance the cultural and political interests and needs of individuals and groups.*					
F4 Resource Use	*Evaluations should use resources effectively and efficiently.*					
P1 Responsive and Inclusive Orientation	*Evaluations should be responsive to stakeholders and their communities.*					
P2 Human Rights and Respect	*Evaluations should be designed and conducted to protect human and legal rights and maintain the dignity of participants and other stakeholders.*					
P4 Clarity and Fairness	*Evaluations should be understandable and fair in addressing stakeholder needs and purposes.*					
P5 Transparency and Disclosure	*Evaluations should provide complete descriptions of findings, limitations, and conclusions to all stakeholders, unless doing so would violate legal and propriety obligations.*					
P6 Conflicts of interests	*Evaluations should openly and honestly identify and address real or perceived conflicts of interests that may compromise the evaluation.*					
P7 Fiscal Responsibility	*Evaluations should account for all expended resources and comply with sound fiscal procedures and processes.*					
A1 Justified Conclusions and Decisions	*Evaluation conclusions and decisions should be explicitly justified in the cultures and contexts where they have consequence.*					
A2 Valid Information	*Evaluation procedures should yield sufficiently dependable and consistent information for the intended uses.*					

A3 Reliable Information	*Evaluation procedures should yield sufficiently dependable and consistent information for the intended use.*					
A4 Explicit Program and Context Descriptions	*Evaluations should document programs and their contexts with appropriate detail and scope for the evaluation purposes.*					
A5 Information Management	*Evaluations should employ systematic information collection, review, verification, and storage methods.*					
A6 Sound Designs and Analyses	*Evaluations should employ technically adequate designs and analyses that are appropriate for the evaluation purposes*					
A7 Explicit Evaluation Reasoning	*Evaluation reasoning leading from information and analyses to findings, interpretations, conclusions, and judgments should be clearly and completely documented.*					
A8 Communication and Reporting	*Evaluation communications should have adequate scope and guard against misconceptions, biases, distortions, and errors.*					
E1 Evaluation Documentation	*Evaluations should fully document their negotiated purposes and implemented designs, procedures, data, and outcomes.*					
E2 Internal Metaevaluation	*Evaluators should use these and other applicable standards to examine the accountability of the evaluation design, procedures employed, information collected, and outcomes.*					
E3 External Metaevaluation	*Program evaluation sponsors, clients, evaluators, and other stakeholders should encourage the conduct of external metaevaluations using these and other applicable standards.*					

NOTE: FRA_yr2_4-Sep-13_Deliverables_Task13_8_22_AttestaFrmAttachTo EvalImpPlan

A.8 Evaluation Standards

The R&D Office's pilot evaluations—and ultimately all of its program evaluations—should be guided by standards of the evaluation profession. This plan recommends initial adoption of the standards contained in Joint Committee on Standards for Educational Evaluation (2011) *The Program Evaluation Standards* (Sage) for use as guiding principles when conducting evaluations. These program evaluation standards are spelled out in the form of 30 standards that are organized in five categories of utility, feasibility, propriety, accuracy, and evaluation accountability. All pilot evaluations should apply the applicable standards and, at an evaluation's end, append to the final report a completed version of the Evaluation Standards Attestation Form which is included in Appendix A.7. Using experience in applying these standards to the pilot evaluations, this plan calls for R&D's program and evaluation personnel to adapt the standards to R&D's special circumstances and needs and subsequently to apply them to meet R&D's ongoing evaluation needs.

a. *Utility standards*. Utility standards are intended to ensure that an evaluation will serve the information needs of the intended users.

Utility standards require evaluations to identify and address stakeholders' needs such that evaluative findings are targeted, scheduled, and delivered to intended users to address their evaluation-related questions and especially to inform their decisions. To meet utility requirements, a program evaluation provides timely, actionable assessments to those persons and groups that are involved in or responsible for implementing the program and to those stakeholders who will use the program's contributions. The person with primary responsibility for conducting the evaluation needs to identify the intended users and their intended uses of findings, then design and carry out the evaluation to provide the users with evaluative feedback that is relevant, clear, concise, and on time. Fully useful evaluations are ones that address the users' most important questions while also obtaining the full range of information required to assess the program's value. The evaluation should not only issue printed reports and other communications of findings, but should also assist users to study and apply the findings. The aim is to do whatever is necessary to secure the evaluation's beneficial impacts. The utility standards reflect the general consensus found in the evaluation literature that program evaluations should effectively address the information needs of clients and other right-to-know audiences, inform program improvement processes, and provide a basis for program accountability. If there is no prospect that the contemplated evaluation's findings would be used, the evaluation should not be undertaken.

b. *Feasibility standards*. Feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

Feasibility standards require those responsible for the evaluation to employ evaluation procedures that are parsimonious and operable in the program's environment. The evaluation should avoid disrupting or otherwise impairing the program. It should control,

as much as possible, the political forces that might otherwise impede or corrupt the evaluation. And it should be conducted as efficiently and cost-effectively as possible. Evaluation procedures must be workable in real-world settings, not only in controlled laboratory settings.

c. ***Propriety standards*. Propriety standards are intended to ensure evaluations will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.**

The propriety standards advise those who will conduct the evaluation to ground the evaluation, from its beginning, in clear, written agreements that define the obligations of the client and evaluator for supporting and executing the evaluation. The evaluation should be designed, executed, and reported to protect all involved parties' rights and dignity. Findings must be honest and not distorted in any way. Reports should be released in accordance with advance editing and disclosure agreements and applicable freedom of information statutes. Moreover, reports should convey appropriately balanced accounts of strengths and weaknesses. The propriety standards reflect the fact that evaluations can affect many people in negative as well as positive ways. The propriety standards are designed to protect the rights of all parties to an evaluation.

d. ***Accuracy standards*. Accuracy standards are intended to ensure that an evaluation will reveal and convey valid and reliable information about all important features of the subject program.**

An evaluation should clearly document the program as it was planned and actually executed. It should describe the program's background and setting and determine its outcomes. It should identify and substantiate the appropriateness of the evaluation's information sources, measurement methods and devices, analytical procedures, and provisions for bias control. It should present the strengths, weaknesses, and limitations of the evaluation's plan, procedures, information, and conclusions. It should describe and assess the extent to which the evaluation provides an independent, unbiased assessment as opposed to a possibly biased self-assessment. In general, this group of standards requires evaluators to obtain technically sound information, analyze it correctly, report justifiable conclusions, and note any pertinent caveats. The overall rating of an evaluation against the accuracy standards is an index of its overall validity.

e. ***Evaluation accountability standards*. Evaluation accountability standards are intended to ensure that those responsible for conducting the evaluation document make available for inspection all aspects of the evaluation that are needed for independent assessments of its utility, feasibility, propriety, accuracy, and accountability.**

The evaluator should document the evaluation's essential details, including how it was planned, how it was executed, what data it obtained, and how it was reported. The evaluator should also make an internal assessment of the evaluation and attest to the extent that it met all of the standards. Additionally, the evaluator should be proactive in seeking, cooperating with, and advocating release of an independent, standards-based

assessment of the evaluation; that is, an external metaevaluation or peer review where needed.