NATIONAL SCIENCE FOUNDATION

4201 WILSON BOULEVARD
ARLINGTON, VIRGINIA 22230

Directorate for Education and Human Resources
Math and Science Partnership (MSP) Program

May 2005

Dear Colleague:

The Math and Science Partnership (MSP) Program at the National Science Foundation is a major Research and Development (R & D) effort designed to integrate the work of higher education with that of K-12 to reform mathematics and science education at all levels, with particular attention to increasing K-12 student achievement. All aspects of the program, including project- and program-level evaluation, are driven by R & D "habits of mind."

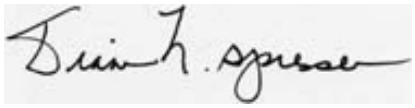
In October 2004, the MSP Program convened a workshop meeting of principal investigators and evaluators of Cohort 1 and 2 projects to formulate a statement that would guide effective project-level evaluation in the context of a national R & D effort, such as the MSP. In recognition of evaluation as an area of expertise and scholarship, the Program sought to bring together this community of evaluators and principal investigators who were experienced in the work of MSP, as well as other experts representing a range of perspectives on evaluation. The Program requested that the leadership of *Building Evaluation Capacity of STEM Projects* -- an MSP-funded project [NSF Grant EHR 0233382] at Utah State University – assume primary responsibility for planning the workshop and for the overall development of any resulting statements and guiding frameworks.

Through the workshop discussions, subsequent discussions by the MSP community at its winter 2005 Learning Network Conference, and a considerable amount of additional work by a team of experienced evaluators, the MSP community has produced the document *Evidence: An Essential Tool – Planning for and Gathering Evidence using the Design-Implementation-Outcomes (DIO) Cycle of Evidence* (NSF 05-31).

On behalf of the MSP program at NSF, I would like to express great appreciation to Drs. Catherine Callow-Heusser, Rosalie T. Torres and Heather J. Chapman, who authored the document, as well as the team of evaluators acknowledged in the introduction to the report. Their expertise, experience and dedication have enabled the MSP program to take an important step forward in project-level evaluation that responds to the ever-deepening nature of an R & D effort. I am especially grateful to Dr. Callow-Heusser for her leadership in this effort. I am also grateful to the broad MSP community of principal investigators and evaluators whose engagement in the development of this document has greatly enhanced its utility.

We ask all MSP Partnership projects to continue their engagement with the *DIO Cycle of Evidence* and to make intelligent use of it as a guiding framework to plan for, gather and use evidence in project-level evaluation. We in the MSP program at NSF also expect to incorporate this framework in the oversight of our projects as we -- and others whom we fund -- engage in the ongoing review of project evaluation commensurate with a major R & D effort.

Sincerely,



Diane M. Spresser
Senior Program Coordinator
National Science Foundation

Evidence: An Essential Tool

Planning for and Gathering Evidence Using the Design-Implementation-Outcomes (DIO) Cycle of Evidence

Prepared for:

The National Science Foundation
Directorate for Education and Human Resources
Math and Science Partnership Program

Prepared by:

The Consortium for Building Evaluation Capacity
2810 Old Main Hill
Utah State University
Logan, UT 84322-2810

April 2005

Evidence: An Essential Tool

Planning for and Gathering Evidence Using the Design-Implementation-Outcomes (DIO) Cycle of Evidence

Prepared under Grant:
EHR-0233382

Catherine Callow-Heusser
Heather J. Chapman
The Consortium for Building Evaluation Capacity

Rosalie T. Torres
Torres Consulting Group

Prepared for The National Science Foundation
Directorate for Education and Human Resources
Math and Science Partnership Program

April 2005

Evidence: An Essential Tool

Evidence: An Essential Tool

Introduction and Background

On July 15, 2004, the Math and Science Partnership (MSP) program at the National Science Foundation (NSF) issued a meeting announcement for "principal investigators and evaluators on MSP Cohort 1 and 2 Partnership and RETA projects to consider and formulate a statement of guidance for effective project-level evaluation in the context of a national R & D effort, such as the MSP" (http://hub.mspnet.org/index.cfm/calendar/show/event-168). Specifically, the MSP Program requested a statement about high quality evidence of effectiveness and efficiency, and a guiding framework that would:

- clarify NSF's expectations for gathering and reporting evidence,
- guide current MSP projects in their evaluation activities, and
- guide future MSPs and other who submit proposals to NSF for funding.

This request arose out of needs to: (a) provide guidance for evaluation planning and evaluation activities to NSF's MSP projects and other projects, and to groups submitting proposals to NSF programs; (b) have a consistent framework by which to assess project-level evaluation; and (c) develop a document about project-level evaluation, grounded in the expertise and experience of the scholarly community having that expertise, that would be an important component of NSF's response to the 2004 Inspector General's Audit of NSF's Math and Science Partnership Program. Both the document and the process by which it was developed are critical components of that response.

The resulting statement and guiding framework—the Design-Implementation-Outcomes (DIO) Cycle of Evidence—described in the following sections were initially outlined by a team of experienced evaluators who have long been involved in evaluations of complex, large-scale projects, particularly mathematics and science projects. This document is based upon work supported by the NSF under supplemental funding to grant EHR-0233382, with the guidance and support of Elizabeth VanderPutten. The developers included (alphabetically)

Catherine Callow-Heusser, Ph.D. Candidate (ABD)
James Dorward, Ph.D., USU's Building Evaluation Capacity MSP-RETA[1]
Joy Frechtling, Ph.D., Westat, consultant for USU's NETA[2]
Frances Lawrenz, Ph.D., U-MN, member of USU's MSP-RETA Advisory Committee
Sean Smith, Ph.D., Horizon Research
Rosalie Torres, Ph.D., Torres Consulting Group, consultant for USU's NETA
Norm Webb, Ph.D., U-WI-Madison

[1]USU: Utah State University; RETA: Research, Evaluation, and Technical Assistance project

[2] NETA: USU's Network for Evaluation Technical Assistance consists of expert evaluators who provide technical assistance to some MSP projects as part of USU's RETA funding

Iris Weiss, Ph.D., Horizon Research, member of USU's MSP-RETA Advisory Committee

Additional input was provided by other members of the MSP-RETA project team at Utah State University (Heather Chapman, Steve Lehman, and Scott Bates), other members of USU's MSP-RETA Advisory Committee (James Altschuld, Frank Davis, Arlen Gullicksen, Donna Mertens, Tom Romberg), and other professional evaluators at Horizon Research. An initial version of the DIO Cycle of Evidence was presented at a meeting of MSP principal investigators (PIs) and evaluators, titled "Evidence: An Essential Tool" and held on October 21-22, 2004, in Arlington, VA. At that meeting, PIs and evaluators reviewed the framework, applied it to activities within their own projects, and provided recommendations for revising and improving the framework. MSP project teams (i.e., PIs, Co-PIs, evaluators, stakeholders) provided additional feedback at the MSP Learning Network Conference held in Washington, DC, January 31-February 1, 2005.

One major theme in the feedback was the need for a common language—a language known to evaluators but also familiar to Science, Technology, Engineering, and Mathematics (STEM) faculty and professionals involved in planning, implementing, and evaluating MSP projects. The DIO Cycle of Evidence addresses this need for a common language, and provides a framework for considering projects and project activities at multiple levels, from the global "big picture" view of projects to the most detailed perspective of individual project activities designed to produce specific outcomes. Overall, the evidence gathered through applying this framework can help increase knowledge, provide evidence for accountability, improve projects, and support the value and feasibility of projects and activities.

The remainder of this document includes the following major sections:

(a) A statement about high quality evidence of project effectiveness and efficiency.

(b) A description of the DIO Cycle of Evidence as a guiding framework for planning, gathering, and using evidence.

(c) The relationship of the DIO Cycle of Evidence to other frameworks used in evaluating projects.

(d) The role of context in establishing evidence of project effectiveness.

(e) Resources to help projects learn more about planning, gathering, and using evidence.

(f) A glossary of terms and abbreviations used throughout this document.

(g) Appendices that contain supplemental resources.

Statement on High Quality Evidence of Effectiveness

The NSF's MSP program is "recognized as a research and development effort for building capacity and integrating the work of higher education - especially its disciplinary faculty in mathematics, the sciences and engineering - with that of K-12 to strengthen and reform science and mathematics education" (NSF, 2005). As with any research and development effort, a focus on high quality evidence of effectiveness and efficiency helps to ensure intellectual rigor and broad impact. To accomplish this, scientifically based evaluation methods for gathering and analyzing evidence must be implemented to determine the effectiveness and efficiency of programs or projects, and to assess the relationship between project implementation and outcomes. Additionally, formative evidence collected along the way should be used to guide ongoing decisions, improve projects and activities, and increase opportunities for successfully attaining project goals.

The MSP program seeks to improve student outcomes in mathematics and science for all K- 12 students. Within the context of MSP, the purpose of evaluation is to provide scientific insights grounded in evidence to

(a) establish the need for MSP projects and activities,
(b) document how the projects are implemented,
(c) improve projects and make data-based decisions about changes for improvements through ongoing formative evaluation, and
(d) determine the impact of projects and activities and demonstrate how impacts were determined.

Using an evaluation framework, and as part of the research and development (R&D) effort integral to the MSP program, MSP partners explore, research, and evaluate methods that best accomplish MSP goals in relation to the five MSP Key Features:

(a) partnerships that effectively engage science, technology, engineering and mathematics (STEM) disciplinary faculty
(b) teacher quality, quantity and diversity
(c) challenging courses and curricula
(d) evidence-based design and outcomes
(e) institutional change and sustainability

High Quality evidence that is both reliable and valid is crucial to determining the degree to which MSP goals are reached. Given the R&D nature of MSPs, methods for gathering the needed evidence must be matched to four main evaluation purposes (Mark, Henry, & Julnes, 2000; Weiss, 1998): (a) oversight and accountability, (b) program improvement involving mid-course corrections, (c) overall assessment of merit and worth, and (d) generating knowledge. Sound evaluation practices, starting with needs assessment (which should be ongoing and continuous), are encouraged to prioritize and conduct evaluation activities to gather evidence for these purposes.

While randomized controlled trials might be best to answer some evaluation questions, most questions within an R&D setting will require alternative or mixed methods (both quantitative and qualitative data gathering and analysis), including interviews, observations, case studies, surveys, and other strategies to understand causality and to provide the information needed to improve educational experiences and outcomes for K-12 students. Applied appropriately, the scientific rigor of these methods can be established. In 2002, in H.R. 3801 included the following definition for scientifically valid educational evaluation:

> The term "scientifically valid education evaluation" means an evaluation that:
> (a) adheres to the highest possible standards of quality with respect to research design and statistical analysis;
> (b) provides an adequate description of the programs evaluated and, to the extent possible, examines the relationship between program implementation and program impacts;
> (c) provides an analysis of the results achieved by the program with respect to its projected effects;
> (d) employs experimental designs using random assignment, when feasible, and other research methodologies that allow for the strongest possible causal inferences when random assignment is not feasible; and
> (e) studies program implementation through a combination of scientifically valid and reliable methods.

The American Evaluation Association (AEA, 2003) proposed that this definition be adopted as it "calls attention to the need for more rigorous methodologies in the context of the function of evaluation to assess and inform. In addition, it illuminates the relationship between program implementation and program impact."[3] Within the MSP, scientifically based evaluation methods for gathering and analyzing evidence can be implemented according to this definition to determine the effectivenss of the MSP program and projects, and to assess the relationship between project implementation and project impact.

While rigorous evaluation designs and appropriate data collection methods are crucial, the evidence gathered and reported forms the foundation for project accountability, project improvement and mid-course corrections, and claims of project impact. *Evidence* is an essential tool for establishing project effectiveness and efficiency.

[3] AEA's statement regarding scientifically based evaluation methods is available at http://www.eval.org/doestatement.htm.

Guiding Framework for Planning, Gathering, and Using Evidence: The Design-Implementation-Outcomes (DIO) Cycle of Evidence

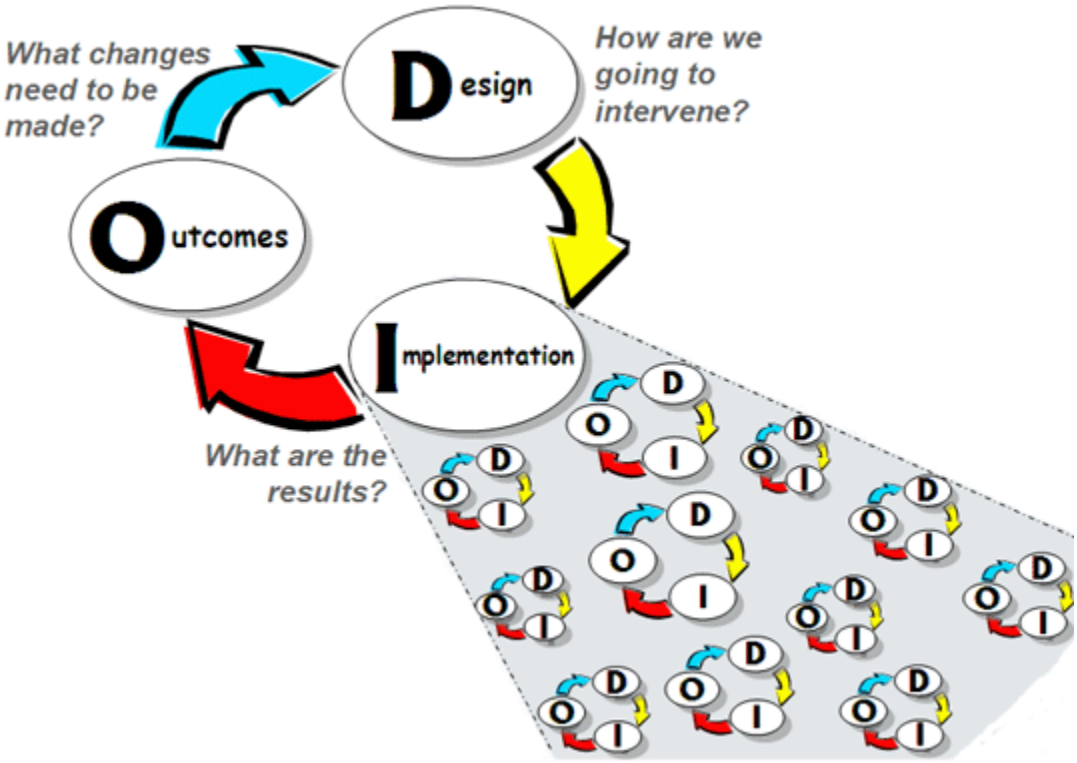
While various research methods, evaluation models, and tools exist to help plan for and gather high quality evidence, high levels of expertise may be required to understand and apply them. Yet, in this time of increased accountability, project designers, implementers, and decision-makers who may be less familiar with evaluation are often responsible for gathering and using evidence throughout the life of the project. The DIO Cycle of Evidence was developed to fill a gap—the gap between experienced evaluators who routinely use evaluation models and tools, and STEM faculty and professionals who are often competent researchers but who have less familiarity with educational evaluation, yet are responsible for evaluations of MSP projects that they plan and implement. The DIO Cycle of Evidence bridges the gap by providing a framework that guides thinking about the design and implementation of evaluation activities, within the context of a research and development cycle and using language not specific to evaluation.

In short, the DIO Cycle of Evidence provides a simple yet rigorous framework for defining the evidence needed to adequately evaluate the design, implementation, and outcomes of a project, or activities conducted within a project. Specifically, it:

- guides planning for and gathering evidence for decision making to help guarantee that MSPs will produce valued outcomes, and to confirm those outcomes.
- prescribes use of valid and reliable evidence to inform the three Design-Implementation-Outcome phases and to make changes and improvements as indicated by the evidence.
- helps prioritize evidence gathering.
- establishes a common language for project personnel, evaluators, and stakeholders to communicate.
- depicts the cyclical nature of designing, implementing, analyzing outcomes, and revising and refining project activities based on those outcomes.
- reminds us that gathering and using evidence is not just an evaluator's responsibility; it is also the responsibility of project designers, implementers, and decision-makers.

The DIO Cycle of Evidence is not an evaluation model. There are many existing evaluation models, and although the DIO Cycle of Evidence phases are congruent with various evaluation models, its main purpose is to guide thinking about, planning for, and gathering evidence. (See the section on "Relationship of the DIO Cycle of Evidence to Other Frameworks" for more information on how the DIO Cycle of Evidence relates to the components of a typical evaluation plan and to logic models.)

As shown in the following figure, the DIO Cycle of Evidence consists of three distinct phases for carrying out and evaluating (a) a project as a whole or (b) specific activities within a project.



While each of the phases will be discussed separately in successive sections, we will point out here that the Implementation phase can involve numerous smaller DIO Cycles of Evidence, with each cycle applied to individual activities within a project, and linked as necessary such that the outcomes of one or more smaller DIO Cycles of Evidence may serve as inputs into other DIO Cycles of Evidence.

Projects may not have sufficient funds to plan for and gather evidence for all of these "mini" DIO Cycles of Evidence, but "key" or major mini-cycles should include evidence-gathering to document effectiveness and to provide evidence that the outcomes of major activities can be attributed to specific preceding activities.

The following pages describe the phases and list specific guiding questions for each phase once the design, activities to be implemented, and outcomes are articulated. Again, this cycle can be applied at many levels—to the overall "global" or "big picture" view of a project, to related groups of activities (e.g., various types of professional development activities all designed to change teacher knowledge and behaviors), or to individual activities.

In all three phases of the DIO Cycle, Design-Implementation-Outcomes, three important questions should be asked along with the questions specific to each phase:

- What EVIDENCE do you need?
- How would you collect the EVIDENCE?
- When does the EVIDENCE need to be collected?

DESIGN

The DESIGN phase is initiated in response to the question, "What problem(s) need to be solved?" Prior to designing a project or activity, a need must be established or problem identified—based for example, on the current state of mathematics and science education. Once a need has been established, the DESIGN phase addresses the overarching question, "How are we going to intervene?"

The DESIGN of a project or activity must be based on evidence that

- Supports the need for the project as well as the design's validity and feasibility,
- Confirms that alternative designs were considered or attempted,
- Indicates that the needs of stakeholders have been addressed, and
- Shows that contextual factors guided decisions about design.

The following questions should guide project or activity DESIGN and can be used as a checklist to provide a comprehensive framework for gathering evidence during the DESIGN phase of a project or activity. To justify the DESIGN of a project or activity, each question should be answered with appropriate evidence supporting the answer.

1. What evidence supports the need for project activities?

- ❑ What are the priority needs and opportunities?
- ❑ What evidence was used to establish these needs or problems?

2. How are we going to intervene?

- ❑ What activities will be planned to bring about change?
- ❑ By what mechanisms (i.e., causal links) will these activities bring about change? What is the program theory or logic model for the design of the project and project activities?
- ❑ Are project activities aligned with project goals? With funding agency goals? With stakeholder goals?
- ❑ Do project activities link to planned outcomes?
- ❑ Is there evidence that the design will solve one or more existing problems or meet established needs?
- ❑ What alternative designs have been tried or could be implemented to address the established need or solve the problem?

3. What evidence is needed to show the design is valid?

- ❑ Is the design supported by relevant research or theory (e.g., published and unpublished reports of studies, descriptions of theory, findings from pilot studies)?
- ❑ If not, how will the validity of the design be justified?
- ❑ What evidence shows that project outcomes have practical importance?
- ❑ Is there evidence that the design will contribute to the knowledge base in meaningful ways?

4. What evidence is needed to show that the needs of different stakeholder groups have been considered?

- ❑ To what extent have the needs of underrepresented groups been considered?
- ❑ To what extent do the values of stakeholder groups differ? How will these differences affect the implementation of project activities?
- ❑ How will differing values affect interpretation of the evidence for the design?
- ❑ What evidence is needed to justify the design to identified audiences or different stakeholder groups? How will the evidence be communicated to various audiences/stakeholders?

5. What evidence is needed to document the contexts within which the project and its activities will operate?
(For additional information, see the section "Role of Context in Establishing Evidence.")

- ❑ What political, social, cultural, or historical factors, values, or characteristics of this setting—including characteristics of schools, teachers, and students—need to be considered?
- ❑ To what extent will contextual factors affect project design, implementation of activities, or outcomes?
- ❑ What alternative designs need to be considered to account for varying contexts?

6. What evidence is needed to show the design is feasible?

- ❑ Can activities based on the project design be accomplished in the given time frame with the given resources (e.g., money, people, skills) within the given context?
- ❑ What changes in the context (e.g., changes in elected officials, policies, other sources of funding, community values, etc.) could affect the feasibility of the design?

Remember, in addition to each question listed above that frames the evidence needed for the DESIGN phase, ask

- How will you collect the evidence?
- When does the evidence need to be collected?

IMPLEMENTATION

The IMPLEMENTATION phase of a project occurs when project activities are carried out in a particular context. Within the IMPLEMENTATION phase, there may be a single DIO Cycle of Evidence or many, potentially interrelated DIO Cycles where the outcomes of one or more DIO Cycles link to subsequent DIO Cycles. Identifying these links and gathering evidence to validate the links are crucial to determining overall project impact. In other words, outcomes of one or more activities may be needed as evidence to establish the design and implementation of other activities, or certain outcomes of one activity may be needed to trigger the start of other activities. The evidence gathered throughout a DIO Cycle of Evidence can impact activities defined by another DIO Cycle, and if outcomes do not meet expectations, the evidence may suggest that mid-course corrections are necessary. These inter-relationships are usually delineated in a project's logic model or theory of action. (See the section on "Relationship of the DIO Cycle of Evidence to Other Frameworks" for more information.)

The IMPLEMENTATION of a project or activity must be based on evidence that

- Demonstrates that project activities have been implemented as planned,
- Explains the degree to which activities were implemented (i.e., implementation fidelity),
- Documents that outcomes of activities were used to guide changes and improvements,
- Confirms that decision-making and mid-course corrections were based on valid data, and
- Identifies contextual factors that could affect implementation in this or other settings.

The following questions should guide project or activity IMPLEMENTATION and can be used as a checklist to provide a comprehensive framework for gathering evidence during the IMPLEMENTATION phase of a project or activity. To justify the IMPLEMENTATION of a project or activity, each question should be answered with appropriate evidence supporting the answer.

1. What evidence is needed to determine if project activities are carried out as planned?

- ❑ Is the project/activity being implemented on schedule? Within budget?
- ❑ What evidence is needed to determine implementation fidelity—the degree to which projects/activities were implemented as planned?
- ❑ How will decisions be made about how much and what kind of evidence to gather (i.e., defining evaluation priorities)?

2. **What evidence is needed to document successes, challenges, and lessons learned? What evidence is needed to document decisions that were made to change implementation of project activities or make mid-course corrections?**

   - ❑ What factors appear to promote successful implementation of project activities?
   - ❑ What barriers hinder implementation of project activities?
   - ❑ What evidence is needed to document deviations to planned implementation? On what evidence were decisions to change planned implementation based? How valid is that evidence?
   - ❑ What lessons have been learned during implementation of project activities?

3. **What evidence is needed to document characteristics of the context, including characteristics of participants, stakeholders, partnerships?**
   (For additional information, see the section "Role of Context in Establishing Evidence.")

   - ❑ Within what contexts were project activities actually implemented? To what extent did these contexts affect implementation?
   - ❑ What are the characteristics of participants, stakeholders, partnerships?
   - ❑ Have contexts/characteristics changed over the course of the project? What accounts for these changes?
   - ❑ What aspects of the context/characteristics might affect outcomes?
   - ❑ Can others use information about implementation of project activities to conduct similar activities to produce similar outcomes in their contexts?

Remember, in addition to each question listed above that frames the evidence needed for the IMPLEMENTATION phase, ask

- How will you collect the evidence?
- When does the evidence need to be collected?

OUTCOMES

The OUTCOMES phase of a project occurs when project activities have been carried out within a particular context and their impact is determined. Within the OUTCOMES phase, data are analyzed to determine (a) if project/activity goals were met, (b) the results or impact of the project or activities, (c) anticipated or unanticipated side effects, and (d) what changes need to be made in DESIGN for successive IMPLEMENTATIONS.

The OUTCOMES of a project or activity must be based on evidence that

- Demonstrates that project or activity goals were reached,
- Shows the extent to which outcomes can be attributed to specific project activities (rather than competing events), and
- Confirms that project/activity outcomes are reliable, valid, cost-effective, and important.

The following questions should guide analysis and interpretation of project or activity OUTCOMES and can be used as a checklist to provide a comprehensive framework for considering evidence during the OUTCOMES phase of a project or activity. To justify the OUTCOMES of a project or activity, each question should be answered with appropriate evidence supporting the answer.

1. **What evidence is needed to determine if anticipated outcomes were achieved?**

   - ❑ In what ways have beginning states been altered in addressing the need or solving the problem? Are these changes sustainable?
   - ❑ What evidence is needed to demonstrate the extent to which outcomes are reliable and valid?
   - ❑ What evidence shows that the needs of underrepresented groups have been addressed?
   - ❑ What evidence establishes sustainable changes in MSP Key Features: teacher quality, quantity, and diversity? Student access to and success in challenging courses and curricula? Partnerships? Changes in higher education and STEM faculty? Institutionalization?

2. **What evidence demonstrates project/activity goals were reached on time and within budget?**

   - ❑ What evidence documents that project/activity goals were reached on schedule and within budget?
   - ❑ What evidence shows why achievement of goals was delayed or not reached?
   - ❑ What evidence documents why achievement of goals resulted in additional expenses?

3. **What evidence is needed to demonstrate the extent to which activity or project outcomes can be attributed to specific project activities?**
   (See the subsequent section on "Logic Models")

   - ❑ What evidence supports the link between project activities and project outcomes?
   - ❑ What competing events or confounding factors could have explained or affected activity/project outcomes?
   - ❑ What gaps or weaknesses in evaluation evidence preclude drawing strong conclusions about the relationships between activities and outcomes?

4. **What aspects of (a) activity or project design, (b) implementation of activities, or (c) evaluation need to be redesigned based on the outcomes, and what evidence is needed to support these changes?**

   - ❑ Were the outcomes expected? Was enough evidence gathered to be able to demonstrate that outcomes were as expected?
   - ❑ If not, what changes will be made to the design and implementation of activities to achieve expected outcomes? What additional evidence needs to be gathered?
   - ❑ Will modifying the existing design help or will a new design be needed?
   - ❑ What evidence will be needed to document changes, improvements, or mid-course corrections?

5. **What evidence is needed to support replication of the project/activities to achieve similar outcomes in other contexts?**
   (For additional information, see the section "Role of Context in Establishing Evidence")

   - ❑ In what ways can the outcomes contribute to the knowledge base?
   - ❑ Can other people use project findings and publications to conduct to conduct similar projects in their contexts? How might outcomes differ in other contexts?

Remember, in addition to each question listed above that frames the evidence needed for the OUTCOMES phase, ask

- How will you collect the evidence?
- When does the evidence need to be collected?

How and When to Gather and Use Evidence

It is not enough to identify what evidence needs to be gathered. When and how evidence will be gathered, as well as how reliability and validity of the evidence will be established, must also be articulated. The following points can guide planning, gathering, analyzing, and reporting evidence.

- Identify the sources of evidence and the requisite time frame for evidence gathering.
- Identify appropriate methods and instruments.
- Identify evidence that exists or is needed to establish reliability and validity of instruments and other devices or processes used to gather evidence.
- Identify how the evidence will be managed and analyzed.
- Plan how the results will be reported and used.

Tables such as those on the following pages can help identify and organize evidence that is needed. The examples show evidence that might be gathered in each phase of the DIO Cycle of Evidence for a new teacher induction and retention model which includes mentoring for new teachers. The activity (as one component of a larger project) and the evidence needed address the MSP Kay Feature, Teacher Quantity, Quality, and Diversity.

Not all sub-questions for the Design, Implementation, and Outcomes phases are included in the following tables, but all main and sub-questions should be considered and their answers justified in planning and gathering evidence. In particular, sub-questions about instrument reliability and validity and quality of evidence are not addressed in the tables, but must be considered in planning, gathering, and analyzing evidence.

This document does not include information about how to analyze evidence or report results. However, USE of results to guide data-based decision-making and support mid-course corrections is an important component of the DIO Cycle of Evidence.

Analysis of evidence for projects as large and complex as MSP projects is beyond the scope of this document. However, the MSP-RETA project at the University of Wisconsin-Madison offers technical assistance to (a) increase the knowledge of design, indicators, and conditions needed to successfully measure change in student learning over time, (b) provide useful tools and designs for evaluators to attribute outcomes to MSP activities, and (c) apply techniques for analyzing the relationship between student achievement and MSP project activities to evaluate the success of MSP projects. Learn more about the "Adding Value to the Mathematics and Science Partnership Evaluations" project at http://www.wcer.wisc.edu/addingvalue/Project%20Mission/Project.htm.

Many resources are available to guide reporting of evaluation results and to make evaluation findings meaningful to stakeholders (see Gangopadhyay, 2002; Miron, 2004; Patton, 1006; Torres, Preskill, & Piontek, 2005).

Evidence Gathering Matrix: DESIGN Phase of a Hypothetical MSP's New Teacher Induction and Mentoring Model

DESIGN Questions	Evidence	Sources	Methods	Timeframe	Instrument(s)
What evidence supports the need for project activities?	Poor retention during first 3 years of teaching	District records	Review of records	Prior to writing MSP proposal	Not Applicable (NA)
	Positive impact of more experienced teachers on student outcomes	Published research	Literature search and review	Prior to writing MSP proposal	Coding sheet of study characteristics*
How are we going to intervene? What evidence is needed to show the design is valid?	Positive impact of new teacher mentoring models on teacher retention leading to more experienced teachers	Published research	Literature search and review	During writing of proposal	Coding sheet of study characteristics*
	Positive impact of the new teacher mentoring model on retention, with less impact (but still positive) of alternative retention model designed to increase content knowledge	Data from pilot test conducted during first year of project funding	Surveys, interviews of new teachers and mentors in pilot district, new teachers in comparison districts	First year of project during which pilot study is implemented	--Surveys, interview protocols developed during pilot study --NAEP Teacher Questionnaire
What evidence is needed to show that the needs of different stakeholder groups have been considered?	Positive impact of mentoring models on retention of teachers from underrepresented groups	Published research	Literature search and review	Prior to writing MSP proposal	Coding sheet of study characteristics*
	Reasons teachers from underrepresented groups exit system	Data from exit interviews	Review of existing data	Prior to writing MSP proposal	NA
	District/school support for implementation of mentoring model	District/school administrators	Letter of support	Prior to writing proposal	NA
What evidence is needed to document the contexts within which the project and its activities will operate?	District, school, teacher, and student characteristics **	--Existing data from district/school records --Existing school culture & climate data	Review of existing data	Prior to writing proposal for MSP funding and during first year of project	NA
What evidence is needed to show the design is feasible?	--Costs in terms of time, personnel expertise, expenses --Context descriptions	Districts who have implemented similar models	Interviews with district administrators	Prior to writing proposal for MSP funding	Interview protocol

* includes reference, study design, sample size, description of intervention, magnitude of impact, rating of study quality, etc.

** including policies about release time for mentoring activities, ethnic percentages of teachers and students, perceptions of administrative support, etc.

Evidence Gathering Matrix: IMPLEMENTATION Phase of a Hypothetical MSP's New Teacher Induction/Mentoring Model

IMPLEMENTATION Questions	Evidence	Sources	Methods	Timeframe	Instrument(s)
What evidence is needed to determine if project activities are carried out as planned?	Changes to project scope and sequence, budgets, personnel	Project documents (e.g., timelines, budget, reports)	Document review	Throughout project	NA
	Number of mentors trained during in-service	Attendance lists for mentor in-service	Document review	Upon completion of in-service	NA
	Activities between mentors, new teachers	Logs of activities between new teachers, mentors	Document review	Throughout school year	Activity log
	Level of support for mentoring model	Administrators, mentors, new teachers	Interview	Every 3 months during school year	Interview protocol
What evidence is needed to document successes, challenges, and lessons learned? What evidence is needed to document decisions that were made to change implementation of project activities or make mid-course corrections?	--Successes --Challenges to implementation, barriers, problems encountered --Problem solutions	--Project documents (e.g., reports, memos, minutes of meetings, email messages, attendance lists) --Teacher mentors and mentees	--Document review --Opinion surveys	--Throughout project and when writing annual/final reports --Annually in spring	--NA --Questionnaires for mentors, mentees
	--Changes to planned activities, timelines, budgets, personnel --Poor outcomes of implemented activities	Project documents and data	--Document review --Data analysis	Throughout project	NA
What evidence is needed to document characteristics of the context, including characteristics of participants, stakeholders, partnerships?	Legislation; state, district, & school policies; district, school, teacher, student, higher ed faculty, & project personnel characteristics; local events	Government documents, district & school records, newspapers, project documents (see above), vitas of project staff & participants	--Document review --Data analysis	Throughout project and when writing annual/final reports	NA
	Characteristics of district/school personnel	District/school administrators, staff, participating teachers	Surveys of contextual characteristics	Beginning and end of school year	e.g., NAEP Teacher Questionnaire
	Characteristics of partnerships	MSP personnel, higher ed faculty, district/school administrators, teachers	Surveys of partnership characteristics	Beginning and end of school year	e.g., instrument developed by MSP-RETA project

Evidence Gathering Matrix: OUTCOMES Phase of a Hypothetical MSP's New Teacher Induction and Mentoring Model

OUTCOMES Questions	Evidence	Sources	Methods	Timeframe	Instrument(s)
What evidence is needed to determine if anticipated project outcomes were achieved? On time and within budget?	Teacher attitudes, characteristics	Mentors, mentees	Opinion surveys	Annually in spring	--Questionnaires --NAEP Teacher Questionnaire
	Changes in teaching practices	Mentees	--Classroom practices survey --Classroom observation	Annually in spring	--Questionnaire --Observation protocol
	Retention data	District records	Document/data review	Annually in fall of successive school years	NA
	Reasons for exiting	Teachers leaving system	Interview	Upon contract non-renewal, leaving position	Exit Interview developed by state department of education prior to MSP
	Impact on student outcomes	Students	Testing	Annually in spring	Standardized or state criterion-referenced tests
	Changes to planned activities, timelines, budgets, personnel	Project documents	--Document review --Data analysis	Throughout project	NA
What evidence is needed to demonstrate the extent to which outcomes can be attributed to specific project activities?	Analysis of other activities/events which may affect outcomes	Documents and data addressing contextual factors	--Document review --Data analysis	End of activity cycle or end of project	NA
What aspects of (a) activity or project design, (b) implementation of activities or (c) evaluation need to be redesigned based on the outcomes, and what evidence is needed to support changes?	--Poor outcomes, poor data --Challenges to implementation --Costs too high (any aspect including personnel) --Activities/outcomes not sustainable --Needs of under-represented group not met, etc.	Project documents, data	Ongoing data analysis	Throughout project	NA
What evidence is needed to support replication of the project/activities to achieve similar outcomes in other contexts?	--Contextual characteristics --Description of implementation	Project documents, data	Final project report, other reports	End of activity cycle or end of project	NA

Relationship of the DIO Cycle of Evidence to Other Frameworks

The DIO Cycle of Evidence does not replace evaluation or R&D models and tools. Rather, it can be used alongside evaluation models as a framework to guide thinking about evidence and to provide additional clarity in planning and gathering evidence for projects. In this section, the relationships between the DIO Cycle of Evidence and other frameworks are described. For more information about evaluation models, including the types of evaluation questions addressed with each model, see Stufflebeam (2001). See also Altschuld & Kumar (2002) for a description of evaluation practices in science and technology.

MSP Key Features

The DIO Cycle of Evidence operationalizes the MSP Key Feature *evidence-based designs and outcomes*. MSP projects are a vehicle designed to initiate change and provide evidence in support of improved student outcomes. *Challenging courses and curricula* and *teacher quality, quantity and diversity* are where the MSP rubber meets the road—these two Key Features are the focus of many MSP project activities. *Partnerships*—another key characteristic—drive the MSP vehicle. Like the confident, experienced driver who makes traversing a difficult, unknown road seem easier, it makes sense that strong partnerships will guide the path and promote success within complex MSP projects. To avoid getting lost or experiencing an unsuccessful journey, the road followed should be selected through *evidence-based designs* that lead to *evidence-based outcomes*, and the turns taken along the way must be supported by decisions that are based on reliable and valid evidence. When funding runs out at the end of the line, *sustainability* resulting from *institutional change* will be needed to keep things on track—to maintain the changes that have occurred. Again, *evidence-based designs and outcomes*, supported by *evidence-based decisions* along the way, will provide the foundation and support needed to sustain changes and impacts attributed to the MSP vehicle.

Activities designed to address *challenging courses and curricula*, to change *teacher quality, quantity and diversity*, to strengthen *partnerships*, and to promote *sustainability* and *institutional change* need to be grounded in *evidence-based designs*, and the impact determined and justified by *evidence-based outcomes*. Additionally, mid-course corrections occurring along the path need to be supported and documented through *evidence-based decision-making*. The DIO Cycle of Evidence guides planning and gathering the evidence needed.

Typical Evaluation Plans

Most evaluation work begins with the development of a plan for carrying out an evaluation study. The format may vary, but plans usually address the following components (Torres, Preskill, & Piontek, 2005):

- Background/organizational context ✓
- Purpose of the evaluation
- Audiences
- Evaluation questions ✓
- Evaluation approach and data collection procedures ✓
- Data analysis procedures
- Evaluation products (including reports to be provided)
- Project management plan (schedule of activities) ✓
- Evaluation constraints
- Budget/costs for the evaluation

The DIO Cycle of Evidence addresses some (see check marks above) but not all of these components. Use of the DIO Cycle of Evidence constitutes a major part of the evaluation planning process, and as such it would be *part* of an overall evaluation plan. Specifically, for its implementation and outcomes phases, the DIO Cycle of Evidence specifies the evaluation questions to be answered; and the data collection (or evidentiary) sources, timeframe, methods, and instruments. This information is virtually the same kind of information that would be found in the evaluation questions, data collection procedures, and project management components of a typical evaluation plan.

Additionally, the Design phase of the DIO Cycle of Evidence provides a rationale and justification for the project's design. Some of this kind of information might be included in the background/organizational context component of an evaluation plan.

Finally, the DIO Cycle of Evidence includes information about *how* the evaluation planning process takes place as well as *how* and *when* decisions based on evaluation findings should be made. In these two aspects, it goes beyond a typical evaluation plan to address both the process of evaluation planning and the use of findings.

Logic Models

Logic models are frequently used to help articulate program theory and to explicate the link between specific project activities and intended project outcomes. However, competently developing and applying logic models requires substantial training and experience, particularly for projects as large and complex as MSP projects.

The DIO Cycle of Evidence was designed to help overcome this need for high levels of expertise—to help articulate the link between project activities and outcomes. Like logic models, the DIO Cycle of Evidence is intended to help produce sound project designs and activities, and to define the evidence required to evaluate designs, activities, and outcomes. While the DIO Cycle of Evidence is not as complex as logic models, it can be subsumed within logic models, and in fact provides specific guidance in the form of checklist questions for better articulating the linkages and evidence needed to attribute outcomes to project activities.

For more information on using logic models, see W. K. Kellogg Foundation (2001); Owen & Rogers (1999); University of WI-Extension (2002).

Role of Context in Establishing Evidence

In a review of evaluation activities conducted on large scale, NSF-funded systemic initiatives, Anderson (2002) reported that (a) the context in which a project was carried out affected interpretation of results, and (b) strategies that were demonstrated to be effective in one context were ineffective in others. Wiess, et al. (2004) distinguishes between "inside" and "outside" forces that effect educational systems. Outside forces include legislation and policies often driven by small but powerful and well-organized groups, and political stances driven by global competitiveness, economic change (e.g., recessions), or disasters that become national or international concerns because of the impact on human lives. Inside forces such as parental concerns, the lack of sufficient numbers of qualified teachers, and children living in poverty also affect educational settings and opportunities for learning. Both inside and outside factors may have large impacts on student outcomes, particularly given the difficulty in controlling and isolating those factors in the typical educational settings in which projects are implemented.

For MSP projects, evidence for contextual factors that have an impact, directly or indirectly, on student outcomes should be gathered and included when analyzing data and interpreting results. Contextual factors include historical, cultural, political, and organizational factors that affect student learning and the environment in which students learn. Some important contextual indicators include student demographics, teacher workload, financial resources, and teacher qualifications (Lashway, 2001), as well as state educational budget changes, testing policies, graduation requirements, and high-impact school or community events (e.g., school closings, threats to safety). The *Program Evaluation Standards* state, "The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified" (Joint Committee, 1994). Examining contextual factors can help explain non-significant or negative findings, unanticipated outcomes, or other unexpected results. Contextual factors can also help guide decisions and provide documentation for mid-course corrections.

Determining which contextual factors to measure can be a daunting task. However, a number of resources are available to help define relevant factors. Some resources are described here.

Key Evaluation Checklist (Scriven, 2004) http://www.wmich.edu/evalctr/checklists/kec.htm

In his Key Evaluation Checklist, Scriven (2004) recommends gathering the following evidence about the background and context of a project:

- Historical, recent, simultaneous, and any projected settings for the program.
- Upstream stakeholders, e.g., NSF, and their stakes.
- Recent relevant legislation and any policy or attitude changes since start-up.
- The underlying rationale, a.k.a. official program theory, and political logic.
- General results of a literature review on similar interventions, including "fugitive" studies not published in standard media and those that can be located on the Internet including the "invisible web" (e.g., by using Copernic Personal Agent).
- Previous evaluations, if any, and their impact.
- Support infrastructure for the project and its activities.

TIMSS Contextual Framework (IEA, 2005) http://timss.bc.edu/timss2003i/context.html
(Schmidt & Cogan, 1996) http://timss.bc.edu/timss1995i/TIMSSPDF/TRCHP5.PDF

The Trends in International Mathematics and Science Study (TIMSS) is designed to help countries all over the world improve student learning in mathematics and science. Educational achievement in approximately 50 countries throughout the world is assessed in the fourth and eighth grades to provide information about trends in performance over time, coupled with extensive background information to address concerns about the quantity, quality, and content of instruction. The TIMSS contextual framework identifies the major characteristics of the educational and social contexts that are studied in relation to improving student learning. The following list categorizes the contextual indicators developed for the TIMSS.

The Curriculum	• Formulating the Curriculum • Scope and Content of the Curriculum • Organization of the Curriculum • Monitoring and Evaluating the Implemented Curriculum • Curricular Materials and Support
The Schools	• School Organization • School Goals • Roles of the School Principal • Resources to Support Mathematics and Science Learning • Parental Involvement • Disciplined School Environment
Teachers and Their Preparation	• Academic Preparation and Certification • Teacher Recruitment, Assignment, and Induction • Teacher Experience • Teaching Styles • Professional Development
Classroom Activities and Characteristics	• Curriculum Topics Taught • Time • Homework • Assessment • Classroom Climate • Information Technology • Calculator Use • Emphasis on Investigation • Class Size
The Students	• Home Background • Prior Experience • Attitudes

NAEP Education Indicators (NCES, 2005) http://nces.ed.gov/programs/coe/list/i4.asp
http://nces.ed.gov/nationsreportcard/mathematics/howdevelop.asp

(CCSSO, 2005) http://www.ccsso.org/naepprofiles/StatebyState/CCD_Master.cfm

The National Assessment of Educational Progress (NAEP), also known as "the Nation's Report Card," is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and the arts. In addition to testing cognitive abilities, NAEP collects information that helps put student achievement in context. Various questionnaires provide context for NAEP assessment results:

- Student questionnaires, which examine background characteristics, subject-area experience, and motivation on the assessment;
- Teacher questionnaires, which gather data on the background and training of teachers and classroom-by-classroom information;
- School questionnaires, which ask principals about school size and other characteristics; and
- SD/LEP questionnaires (students with disabilities or limited English proficiency), which ask the person most familiar with the student about accommodations normally permitted.

Resources for Planning, Gathering, and Using Evidence

This section briefly presents information on the characteristics of high quality evaluation, as a general practice; and on the characteristics of high quality evidence, in particular. It concludes with a listing of online resources for planning, gathering, and using evidence.

Standards for High Quality Evaluation Practice

Most evaluation practice is guided by two sets of guidelines established within the profession: The Program Evaluation Standards of the Joint Committee on Standards for Educational Evaluation, and the Guiding Principles of the American Evaluation Association (AEA).

Joint Committee's Program Evaluation Standards

The Joint Committee on Standards for Educational Evaluation was founded in 1975 to develop standards for educational evaluation. Originally initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, the Joint Committee now includes many other organizations in its membership. The Joint Committee has developed standards for evaluating educational programs as well as for evaluating personnel.[4]

A summary of the Program Evaluation Standards follow and can be found at http://www.eval.org/EvaluationDocuments/progeval.html (see also Appendix A). The full text of *The Program Evaluation Standards* (2nd edition) is available for purchase from Sage Publications at http://www.sagepub.com/. Additionally, see "What the Program Evaluation Standards Say about Designing Evaluations," available at http://www.wmich.edu/evalctr/jc/DesigningEval.htm, and NCREL (2005).

Utility Standards. The utility standards are intended to ensure that an evaluation will serve the information needs of intended users. Utility standards include the following:

U1 Stakeholder Identification
U2 Evaluator Credibility
U3 Information Scope and Selection
U4 Values Identification
U5 Report Clarity
U6 Report Timeliness and Dissemination
U7 Evaluation Impact

[4] Further information about the Joint Committee's work and reprint requests may be addressed to: The Joint Committee on Standards for Educational Evaluation, The Evaluation Center, Western Michigan University, Kalamazoo MI 49008-5178.

Feasibility Standards. The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal. Feasibility standards include the following:

F1 Practical Procedures
F2 Political Viability
F3 Cost Effectiveness

Propriety Standards. The propriety standards are intended to ensure that an evaluation 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. Propriety standards include the following:

P1 Service Orientation
P2 Formal Agreements
P3 Rights of Human Subjects
P4 Human Interactions
P5 Complete and Fair Assessment
P6 Disclosure of Findings
P7 Conflict of Interest
P8 Fiscal Responsibility

Accuracy Standards. The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated. Accuracy standards include the following:

A1 Program Documentation
A2 Context Analysis
A3 Described Purposes, Procedures
A4 Defensible Information Sources
A5 Valid Information
A6 Reliable Information
A7 Systematic Information
A8 Analysis of Quantitative Information
A9 Analysis of Qualitative Information
A10 Justified Conclusions
A11 Impartial Reporting
A12 Metaevaluation

American Evaluation Association's Guiding Principles

In 1994, the membership of AEA adopted a set of principles developed to guide the professional practice of evaluators, and to inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators. A revision of the 1994 Guiding Principles was ratified by AEA membership in 2004. A summary of the Guiding Principles follows. The principles can be found in complete detail along with additional information about the development process at http://www.eval.org/Guiding%20Principles.htm. See also Appendix B.

(a) **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries about whatever is being evaluated.

(b) **Competence:** Evaluators provide competent performance to stakeholders.

(c) **Integrity/Honesty:** Evaluators ensure the honesty and integrity of the entire evaluation process.

(d) **Respect for People:** Evaluators respect the security, dignity, and self-worth of the respondents, program participants, clients, and other stakeholders with whom they interact.

(e) **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of interests and values that may be related to the general and public welfare.

Characteristics of High Quality Evidence

Both the *Program Evaluation Standards* and the *Guiding Principles for Evaluators* speak to evaluators' use of high quality evidence. In particular, the Systematic Inquiry principle of the Guiding Principles calls for evaluators to "adhere to the highest technical standards appropriate to the methods they use"; the Accuracy Standards address validity, reliability, and the analysis of both quantitative and qualitative data.

Quantitative methods for gathering evidence are judged by two main criteria: reliability and validity. Reliability forms the answer to the question, "Were our measurements consistent?" Validity answers the question, "Did we measure what we were supposed to measure?" Definitions for reliability and validity [adapted from The Evaluation Center's Glossary (Wheeler, Haertel, & Scriven, 1992); available at http://ec.wmich.edu/glossary/index.htm] are as follows:

Reliability: the degree to which an assessment or instrument consistently measures an attribute. There are several types of reliabilities, for example:

- **Intra-Rater** - the degree to which the measure yields consistent results for the same individual over different administrations.
- **Inter-Rater** - the degree to which the measure yields similar results when multiple assessors use the same instrument to measure an individual at a given point in time.
- **Internal Consistency** - the degree to which individual observations or items consistently measure the same attribute.
- **Test-Retest** - the degree to which the measure produces consistent results over several administrations assessing the same attribute of a teacher.

Validity: the extent to which the test scores or responses measure the attribute(s) that they were designed to measure. Several types of validity are described below:

- **Concurrent** - the relationship of one measure to another simultaneous measure or variable assessing the same or a related attribute.
- **Construct** - the degree of fit of a measure and its interpretation with its underlying explanatory concepts, theoretical rationales, or foundations.
- **Content** - (1) the appropriateness of the domain definition and the sampling of content; (2) the extent of congruence between the scope of a content area that an instrument or process claims to cover and what it actually does cover. Both definitions are aspects of construct validity.
- **Criterion-Related** - the correlation or extent of agreement of the test score from an assessment with one or more external variables that measure the attribute being assessed.
- **Curricular** - the extent to which the items on the assessment or test measure the content of a local curriculum, or the extent of agreement between the test coverage (topics, breadth and depth, skills, cognitive complexity) and the goals and objectives of the curriculum.
- **Instructional** - the degree to which the items on a test measure: (a) what is actually being taught, and (b) what the individuals being assessed have had an opportunity to learn.
- **Face** - the perceived extent of acceptability or legitimacy of an instrument or process to teachers, administrators, policymakers, students, parents, the general public, and other stakeholders.

Qualitative methods for gathering evidence are judged by criteria different than quantitative data, as shown in the following table adapted from The Research Methods Knowledge Base (Trochim, 2002; available at http://www.socialresearchmethods.net/kb/).

Criteria for Judging Quantitative Data	Criteria for Judging Qualitative Data
Validity	Credibility Transferability
Reliability	Dependability
Objectivity	Confirmability

Definitions of the criteria for judging qualitative data (adapted from The Research Methods Knowledge Base, Trochim, 2002; see also Golafshani, 2003) are as follows:

Credibility: involves establishing that the results of qualitative research are credible or believable from the perspective of the participant in the research or the decision-maker using the findings of the research.

Transferability: refers to the degree to which the results of qualitative research can be generalized or transferred to other contexts or settings. From a qualitative perspective transferability is primarily the responsibility of the one doing the generalizing. The qualitative researcher can enhance transferability by doing a thorough job of describing the research context and the assumptions that were central to the research.

Dependability: emphasizes the need for the researcher to account for the ever-changing context within which research occurs. The researcher is responsible for describing the changes that occur in the setting and how these changes affected the way the study was approached.

Confirmability: refers to the degree to which the results could be confirmed or corroborated by others. There are a number of strategies for enhancing confirmability. The researcher can document the procedures for checking and rechecking the data throughout the study, and after the study, can conduct a data audit to examine the data collection and analysis procedures and makes judgments about the potential for bias or distortion.

The DIO Cycled of Evidence promotes planning and gathering high quality evidence, as judged by the preceding criteria for quantitative and qualitative data. Evidence defined by the checklist questions associated with the phases of the DIO Cycle of Evidence (see the section, Guiding Framework for Planning, Gathering, and Using Evidence: The Design-Implementation-Outcomes Cycle of Evidence) includes evidence to support the reliability and validity of measures and the evidence gathered using them.

Online Resources for Planning, Gathering, and Using Evidence

Many resources for planning, gathering, and using evidence exist, and while the authors of this document do not promote any one over others, the following websites for educational evaluation and research associations, NSF-funded projects, and NSF-published evaluation documents can help you learn more and locate resources relevant to your project's evaluation needs.

American Evaluation Association (AEA) http://www.eval.org

The American Evaluation Association is "devoted to the application and exploration of evaluation in all its forms." AEA's webpage is a great resource for all evaluators and others needing to learn more about evaluation or conduct evaluations. Complete with links to other evaluation resources as well as lists of members and current topics in evaluation, this site has information for evaluators in all fields.

AEA's Evaluation Links http://eval.org/EvaluationLinks/

AEA's evaluation links include a variety of information resources such as professional groups affiliated with evaluation, evaluation consultants, electronic discussion groups, resources for high-stakes testing, links to qualitative data analysis software and survey design, administration, scanning, and analysis products, AEA's statement about scientifically-based evaluation methods, and a "Collection of Links" page that offers links to different web-based resource pages dealing with a variety of evaluation topics. The links include the following:

- The Evaluation Clearinghouse
- On-Line Evaluation Resource Library
- Resources for Methods in Evaluation and Social Research
- The WWW Virtual Library: Evaluation

AEA's Link to Online Texts http://www.eval.org/EvaluationLinks/onlinehbtxt.htm

The "Online Handbooks and Texts" page accessible through the AEA website offers links to 30+ handbooks and texts available online in their entirety. Subjects include but are not limited to designing evaluations, choosing the proper methodological techniques, assessing impact, and proper use and interpretation of statistical methods.

American Educational Research Association (AERA) http://www.aera.net/

The American Educational Research Association (AERA) is concerned with improving the educational process by encouraging scholarly inquiry related to education and by promoting the dissemination and practical application of research results. AERA is the most prominent international professional organization with the primary goal of advancing educational research and its practical application. The association's website offers links to journals and textbooks related to educational research. Textbooks can be purchased through this link, and journal access is available to members. Divisions focusing on broad substantive or professional interests, including the following:

- Division B: Curriculum Studies
- Division C: Learning & Instruction
- Division D: Measurement & Research Methodology
- Division G: Social Context of Education
- Division H: School Evaluation & Program Development
- Division J: Postsecondary Education
- Division K: Teaching & Teacher Education
- Division L: Educational Policy & Politics

The Evaluation Center http://www.wmich.edu/evalctr/

The Evaluation Center's mission is to advance the theory, practice, and utilization of evaluation. The Center's principal activities are research, development, dissemination, service, instruction, and national and international leadership in evaluation. This site offers access to evaluation checklists, journals, a directory of evaluators, and much more.

The Evaluation Center's Evaluation Checklists http://www.wmich.edu/evalctr/checklists/

This link offers access to checklists developed by top evaluators in the field and funded by NSF. Examples include the following among many checklists. The site is updated frequently.

- The Key Evaluation Checklist (Scriven, 2005)
- The Evaluation Design Checklist (Stufflebeam, 2004),
- The Evaluation Reports Checklist (Miron, 2004)
- A Checklist for Evaluating Large-Scale Assessment Programs (Shepard, 1977)
- Making Evaluation Meaningful to all Education Stakeholders (Gangopadhyay, 2002)
- Utilization-Focused Evaluation (Patton, 2002)

SRI's On-Line Evaluation Resource Library (OERL) http://oerl.sri.com/

SRI's On-Line Evaluation Resource Library was funded by NSF and developed for professionals seeking to design, conduct, document, or review project evaluations. Its mission is to support the continuous improvement of project evaluations. Specific examples of evaluation plans for the following areas are available: curriculum development, teacher education, faculty development, laboratory development, under-represented populations, and technology.

National Science Foundation (NSF) http://www.nsf.gov/

The NSF website includes administrative details pertinent to any NSF funded project, such as links to general information, staff directory, upcoming events, discoveries of NSF research and more. The following publications related to evaluation are available:

- The Cultural Context of Educational Evaluation: The Role of Minority Evaluation Professionals, http://www.nsf.gov/pubs/2001/nsf0143/start.htm
- FOOTPRINTS: Strategies for Non-Traditional Program Evaluation http://www.nsf.gov/pubs/1995/nsf9541/nsf9541.pdf
- The 2002 User-Friendly Handbook for Project Evaluation http://www.nsf.gov/pubs/2002/nsf02057/start.htm
- The Cultural Context of Educational Evaluation: A Native American Perspective http://www.nsf.gov/pubs/2003/nsf03032/start.htm
- User-Friendly Handbook for Mixed Method Evaluations http://www.nsf.gov/pubs/1997/nsf97153/start.htm

Glossary

Abbreviations

AEA	American Evaluation Association, http://www.eval.org The AEA is an international professional association of evaluators devoted to the application and exploration of program evaluation, personnel evaluation, technology, and many other forms of evaluation. Evaluation involves assessing the strengths and weaknesses of programs, policies, personnel, products, and organizations to improve their effectiveness.
AERA	American Educational Research Association, http://www.aera.net The AERA, founded in 1916 and currently in its 86th year, is concerned with improving the educational process by encouraging scholarly inquiry related to education and by promoting the dissemination and practical application of research results. See also the Divisions within AERA that focus on broad substantive or professional interests, ranging from administration and curriculum to teacher education and education policy and politics.
CCSSO	Council of Chief State School Officers, http://www.ccsso.org
DIO	Design-Implementation-Outcomes Cycle of Evidence
IAE	International Association for the Evaluation of Educational Achievement
MSP	With the Math and Science Partnership Program, launched in 2002, the NSF assumed important responsibilities for building the capacity to implement a key facet of the No Child Left Behind (NCLB) vision for K-12 education. The MSP program includes 4 kinds of projects: Comprehensive, Targeted, RETA, and Institute Partnerships.
MSP-NET	The Math and Science Partnership's Learning Network, http://hub.mspnet.org/
NAEP	The National Assessment of Educational Progress (NAEP), also known as "the Nation's Report Card," is the only nationally representative and continuing assessment of what America's students know and can do in various subject areas. Since 1969, assessments have been conducted periodically in reading, mathematics, science, writing, U.S. history, civics, geography, and the arts.
NETA	Utah State University's Network for Evaluation Technical Assistance, consisting of expert evaluation consultants who provide technical assistance to some MSP projects through USU's RETA project
NSF	The National Science Foundation was created by Congress in 1950 "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense…" http://www.nsf.gov

PI	Principal Investigator
R&D	Research and development
RETA	NSF's MSP Research, Evaluation, and Technical Assistance projects
STEM	Science, Technology, Engineering, and Mathematics
TIMMS	The Trends in International Mathematics and Science Study is designed to help countries all over the world improve student learning in mathematics and science. Educational achievement in about 50 countries throughout the world is assessed in the fourth and eighth grades to provide information about trends in performance over time together with extensive background information to address concerns about the quantity, quality, and content of instruction.
USU	Utah State University

Definitions of Terms

Context	The specific setting a program occurs in. This includes social, political, cultural, historical, and personal factors.
External Validity	The extent to which evaluation findings are generalizable to other cases, settings, contexts, and times.
Feasibility of the Evaluation	Extent to which the evaluation is realistic, prudent, diplomatic, and frugal given the context within which it occurs. In order to claim feasibility, an evaluation plan should use practical procedures, consider the positions of differing stakeholder groups, and be efficient with a goal of providing information of sufficient value to justify the process.
Impact Theory	The nature of the change in social conditions brought about by program action
Implementation Fidelity	The extent to which a program or intervention has been implemented in a particular setting. In particular, the degree to which the actual implementation differs from the planned implementation.
Internal Validity	The extent to which contextual factors affect the relationship between the program (or an intervention) and outcomes. When we assess the degree to which contextual factors affect the relationship between the program and its outcomes, we are able to better determine the level of confidence with which we can attribute outcomes to the program or intervention.

Measurement Validity	Extent to which an instrument measures what it is intended to measure. A valid measure yields results consistent with past work using the same concept; is consistent with alternative measures that have been used to assess the same concept; is internally consistent; has an adequate degree of predictability. See definitions in section, "Characteristics of High Quality Evidence."
Meta-analysis	The systematic analysis of the results of a body of evaluations and research studies of similar programs to produce an estimate of overall program effect, determine conditions under which better outcomes are realized, and examine the characteristics of evaluations that influence the kind of effects found.
Need	Discrepancy between a set of existing versus desired conditions a program is intended to address.
Needs Assessment	A systematic set of procedures undertaken for the purpose of setting priorities and making decisions about program or organizational improvement and allocation of resources. The priorities are based upon identified needs. (Witkin & Altschuld, 1995)
Pilot Study	Small, preliminary test or trial run of an intervention, or of an evaluation activity such as an instrument or sampling procedure. The results of the pilot are used to improve the program or evaluation procedure being piloted before it is used on a larger scale.
Program Theory	The set of assumptions about the manner in which the program relates to the social benefits it is expected to produce, and the strategy and tactics the program has adopted to achieve its goals and objectives.
Reliability	The extent to which scores obtained on a measure are reproducible in repeated administrations under the same measurement conditions; results that fail to prove reliable risk underestimation of effect. See definition in section, "Characteristics of High Quality Evidence."
Stakeholders	Any person legitimately involved in or affected by a project or its evaluation, including students, their parents/guardians, teachers, other school or district staff, superintendents, state legislators, future employers, funding agencies, and others who make decisions about or are affected by a project or its evaluation.
Theory	An explanation of the commonalities and the relationships among observed phenomena in terms of the causal structures and processes that are presumed to underlie them.

Utility of the Evaluation	The extent to which an evaluation will serve the information needs of, or be valuable to, intended users.
Validity	The extent to which the construct under investigation was measured—"Did we measure what we thought we were measuring? See definition in section, "Characteristics of High Quality Evidence."

For additional information, see Cousins & Earl (1995); Cousins & Whitmore (1998); Fitzpatrick, Sanders, & Worthen (2004); Gall, Gall, & Borg (2003); Mark, Henry, & Julnes (2000); Mathison (2005); Patton (1996); Rossi, Lipsey, & Freeman (2004); Stufflebeam (2001); Torres, Preskill, & Piontek (2005); Trochim (2004); Weiss (1998); Witkin & Altschuld (1995).

References

Altschuld, J. W., & Kumar, D. D. (2002). *Evaluation of science and technology education at the dawn of a new millennium*. New York, NY: Kluwer Academic/Plenum Publishers.

AEA. (2003). *American Evaluation Association response to U. S. Department of Education Notice of proposed priority, Federal Register RIN 1890-ZA00, November 4, 2003, "Scientifically Based Evaluation Methods*". Last retrieved on March 20, 2005, at http://www.eval.org/doestatement.htm

Anderson, B. (2002). Evaluating systemic reform: Evaluation needs, practices, and challenges. In J. W. Altschuld & D. D. Kumar (eds.), *Evaluation of science and technology education at the dawn of a new millennium*, pp. 49-80. New York, NY: Kluwer Academic/Plenum Publishers.

CCSSO. (2005). *Profiles of state education systems for use with NAEP*. Last retrieved on March 20, 2005, from http://www.ccsso.org/naepprofiles/StatebyState/CCD_Master.cfm

Cousins, J. B., & Earl, L. M., Editors. (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer.

Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80*, 5-24. San Francisco, CA: Jossey Bass.

Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines, 3rd edition.* Boston, MA: Allyn and Bacon.

Gall, M. D., Gall, J. P. & Borg, W. R. (2003). *Educational research: An introduction*. Boston, MA: A B Longman.

Gangopadhyay, (2002). *Making evaluation meaningful to all education stakeholders*. Last retrieved on March 20, 2005, from http://www.wmich.edu/evalctr/checklists/makingevalmeaningful.htm

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report, 8*(4), 597-607. Available at http://www.nova.edu/ssss/QR/QR8-4/golafshani.pdf

IEA. (2005). *TIMMS contextual background questionnaires*. Last retrieved on March 20, 2005, from http://timss.bc.edu/timss2003i/context.html

Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards, 2nd edition*. Thousand Oaks, CA: Sage Publications.

Lashway, L. (2001). *Educational indicators*. ERIC Digest. ED457536 available at http://www.ericdigests.org/2002-2/indicators.htm

Mark, M. M., Henry, G. T., & Julnes, G. (2000). *Evaluation: An integrated framework for understanding, guiding, and improving policies and programs*. San Francisco, CA: Jossey Bass.

Mathison, S. (Ed.) (2005). *Encyclopedia of evaluation*. Thousand Oaks, CA: Sage Publications.

Miron, G. (2004). *Evaluation report checklist*. Last retrieved on March 20, 2005, from http://www.wmich.edu/evalctr/checklists/reports.pdf

NCES. (2005). *How the NAEP Mathematics Assessment Works*. Last retrieved on March 20, 2005, from http://nces.ed.gov/nationsreportcard/mathematics/

NCREL. (2005). *Evaluating progress: More on evaluation standards*. Last retrieved on February 18, 2005, from http://www.ncrel.org/tandl/eval6.htm

NSF. (1996). The value of evaluation. In *NSF Evaluation Highlights: A Report on the National Science Foundation's Efforts to Assess the Effectiveness of Its Education Programs*, pp. 1-10. Washington, DC: Author.

NSF. (2005). *Math and science partnership program*. Last retrieved from http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5756&org=EHR&from=fund on March 20, 2005.

Owen, J. & Rogers, P. J. (1999). *Program evaluation: Forms and approaches*. Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (1996). *Utilization-focused evaluation : The new century text (3rd edition).* Thousand Oaks, CA: Sage Publications.

Patton, M. Q. (2002). *Utilization-focused evaluation (UFE) checklist*. Last retrieved on March 20, 2005, from http://www.wmich.edu/evalctr/checklists/ufechecklist.htm

Rossi, P. H., Lipsey, M. W. & Freeman, H. E. (2004). *Evaluation: A systematic approach*, 7th edition. Thousand Oaks, CA: Sage Publications.

Schmidt, W.H. & Cogan, L.S. (1996). Development of the TIMSS context questionnaires" in M.O. Martin & D.L. Kelly (eds.), *Third International Mathematics and Science Study (TIMSS) Technical Report, Volume I: Design and Development*. Chestnut Hill, MA: Boston College.

Scriven, M. (2005). *Key evaluation checklist*. Last retrieved from http://www.wmich.edu/evalctr/checklists/kec.htm on March 20, 2005.

Shepard, L. A. (1977). *A checklist for evaluating large-scale assessment programs*. Last retrieved on March 20, 2005, from http://www.wmich.edu/evalctr/checklists/assessment_eval_introframe.htm

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation, 89*, 7-98. San Francisco, CA: Jossey Bass.

Stufflebeam, D. L. (2004). *Evaluation design checklist.* Last retrieved on March 20, 2005, from http://www.wmich.edu/evalctr/checklists/evaldesign.htm

Torres, R. T., Preskill, H., & Piontek, M. (2005). *Evaluation strategies for communicating and reporting : Enhancing learning in organizations*, 2nd ed. Thousand Oaks, CA: Sage Publications.

Trochim, W. M. (2004). *The research methods knowledge base, 2nd edition*. Internet WWW page, at URL: <http://trochim.human.cornell.edu/kb/index.htm> (version current as of August 16, 2004).

University of Wisconsin-Extension. (2002). *Enhancing program performance with logic models.* Internet WWW page, at URL:< http://www.uwex.edu/ces/lmcourse/>.

Weiss, C. H. (1998). *Evaluation,* 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Weiss, I. R., Knapp, M. S., Hollweg, K. S., & Burrill, G., (eds). (2004). *Investigating the influence of standards: A framework for research in mathematics, science, and technology education.* Washington, DC: National Academies Press.

Wheeler, P., Haertel, G. D., & Scriven, M. (1992). *Teacher evaluation glossary*. Kalamazoo, MI: CREATE Project, The Evaluation Center, Western Michigan University

Witkin, B. R., & Altschuld, J. W. (1995). *Planning and conducting needs assessments : A practical guide.* Thousand Oaks, CA: Sage Publications.

W. K. Kellogg Foundation (2001). *Logic model development guide*. Battle Creek, MI: Author. (http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf)

Appendix

Program Evaluation Standards

The Joint Committee on Standards for Educational Evaluation was founded in 1975 to develop standards for educational evaluation. Originally initiated by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, the Joint Committee now includes many other organizations in its membership. The Joint Committee has developed a set of standards for the evaluation of educational programs as well as for evaluating personnel. Further information about the Joint Committee's work and reprint requests may be addressed to: The Joint Committee on Standards for Educational Evaluation, The Evaluation Center, Western Michigan University, Kalamazoo MI 49008-5178.

The summary of the Program Evaluation Standards that follows can be found at http://www.eval.org/EvaluationDocuments/progeval.html .

The full text of the Program Evaluation Standards (2nd edition) can be purchased through Sage Publications at http://www.sagepub.com/.

Additionally, see "What the Program Evaluation Standards Say about Designing Evaluations," available at http://www.wmich.edu/evalctr/jc/DesigningEval.htm.

Summary of the Standards

Utility Standards

The utility standards are intended to ensure that an evaluation will serve the information needs of intended users.

U1. **Stakeholder Identification:** Persons involved in or affected by the evaluation should be identified, so that their needs can be addressed.

U2. **Evaluator Credibility:** The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that the evaluation findings achieve maximum credibility and acceptance.

U3. **Information Scope and Selection:** Information collected should be broadly selected to address pertinent questions about the program and be responsive to the needs and interests of clients and other specified stakeholders.

U4. **Values Identification:** The perspectives, procedures, and rationale used to interpret the findings should be carefully described, so that the bases for value judgments are clear.

U5. **Report Clarity:** Evaluation reports should clearly describe the program being evaluated, including its context, and the purposes, procedures, and findings of the evaluation, so that essential information is provided and easily understood.

U6. **Report Timeliness and Dissemination:** Significant interim findings and evaluation reports should be disseminated to intended users, so that they can be used in a timely fashion.

U7. **Evaluation Impact:** Evaluations should be planned, conducted, and reported in ways that encourage follow-through by stakeholders, so that the likelihood that the evaluation will be used is increased.

Feasibility Standards

The feasibility standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.

F1. **Practical Procedures:** The evaluation procedures should be practical, to keep disruption to a minimum while needed information is obtained.

F2. **Political Viability:** The evaluation should be planned and conducted with anticipation of the different positions of various interest groups, so that their cooperation may be obtained, and so that possible attempts by any of these groups to curtail evaluation operations or to bias or misapply the results can be averted or counteracted.

F3. **Cost Effectiveness:** The evaluation should be efficient and produce information of sufficient value, so that the resources expended can be justified.

Propriety Standards

The propriety standards are intended to ensure that an evaluation 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.

P1. **Service Orientation:** Evaluations should be designed to assist organizations to address and effectively serve the needs of the full range of targeted participants.

P2. **Formal Agreements:** Obligations of the formal parties to an evaluation (what is to be done, how, by whom, when) should be agreed to in writing, so that these parties are obligated to adhere to all conditions of the agreement or formally to renegotiate it.

P3. **Rights of Human Subjects:** Evaluations should be designed and conducted to respect and protect the rights and welfare of human subjects.

P4. **Human Interactions:** Evaluators should respect human dignity and worth in their interactions with other persons associated with an evaluation, so that participants are not threatened or harmed.

P5. **Complete and Fair Assessment:** The evaluation should be complete and fair in its examination and recording of strengths and weaknesses of the program being evaluated, so that strengths can be built upon and problem areas addressed.

P6. **Disclosure of Findings:** The formal parties to an evaluation should ensure that the full set of evaluation findings along with pertinent limitations are made accessible to the persons affected by the evaluation, and any others with expressed legal rights to receive the results.

P7. **Conflict of Interest:** Conflict of interest should be dealt with openly and honestly, so that it does not compromise the evaluation processes and results.

P8. **Fiscal Responsibility:** The evaluator's allocation and expenditure of resources should reflect sound accountability procedures and otherwise be prudent and ethically responsible, so that expenditures are accounted for and appropriate.

Accuracy Standards

The accuracy standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features that determine worth or merit of the program being evaluated.

A1. **Program Documentation:** The program being evaluated should be described and documented clearly and accurately, so that the program is clearly identified.

A2. **Context Analysis:** The context in which the program exists should be examined in enough detail, so that its likely influences on the program can be identified.

A3. **Described Purposes and Procedures:** The purposes and procedures of the evaluation should be monitored and described in enough detail, so that they can be identified and assessed.

A4. **Defensible Information Sources:** The sources of information used in a program evaluation should be described in enough detail, so that the adequacy of the information can be assessed.

A5. **Valid Information:** The information gathering procedures should be chosen or developed and then implemented so that they will assure that the interpretation arrived at is valid for the intended use.

A6. **Reliable Information:** The information gathering procedures should be chosen or developed and then implemented so that they will assure that the information obtained is sufficiently reliable for the intended use.

A7. **Systematic Information:** The information collected, processed, and reported in an evaluation should be systematically reviewed and any errors found should be corrected.

A8. **Analysis of Quantitative Information:** Quantitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A9. **Analysis of Qualitative Information:** Qualitative information in an evaluation should be appropriately and systematically analyzed so that evaluation questions are effectively answered.

A10. **Justified Conclusions:** The conclusions reached in an evaluation should be explicitly justified, so that stakeholders can assess them.

A11. **Impartial Reporting:** Reporting procedures should guard against distortion caused by personal feelings and biases of any party to the evaluation, so that evaluation reports fairly reflect the evaluation findings.

A12. **Metaevaluation:** The evaluation itself should be formatively and summatively evaluated against these and other pertinent standards, so that its conduct is appropriately guided and, on completion, stakeholders can closely examine its strengths and weaknesses.

American Evaluation Association Guiding Principles for Evaluators

In 1994 the membership of AEA adopted a set of principles developed to guide the professional practice of evaluators, and inform evaluation clients and the general public about the principles they can expect to be upheld by professional evaluators. A revision of the 1994 Guiding Principles was ratified by AEA membership in 2004. A summary of the Guiding Principles follows. Along with additional information about the development process, the guidelines as shown below can be found at http://www.eval.org/Guiding%20Principles.htm;

A. **Systematic Inquiry:** Evaluators conduct systematic, data-based inquiries.

1. To ensure the accuracy and credibility of the evaluative information they produce, evaluators should adhere to the highest technical standards appropriate to the methods they use.

2. Evaluators should explore with the client the shortcomings and strengths both of the various evaluation questions and the various approaches that might be used for answering those questions.

3. Evaluators should communicate their methods and approaches accurately and in sufficient detail to allow others to understand, interpret and critique their work. They should make clear the limitations of an evaluation and its results. Evaluators should discuss in a contextually appropriate way those values, assumptions, theories, methods, results, and analyses significantly affecting the interpretation of the evaluative findings. These statements apply to all aspects of the evaluation, from its initial conceptualization to the eventual use of findings.

B. **Competence:** Evaluators provide competent performance to stakeholders.

1. Evaluators should possess (or ensure that the evaluation team possesses) the education, abilities, skills and experience appropriate to undertake the tasks proposed in the evaluation.

2. To ensure recognition, accurate interpretation and respect for diversity, evaluators should ensure that the members of the evaluation team collectively demonstrate cultural competence. Cultural competence would be reflected in evaluators seeking awareness of their own culturally-based assumptions, their understanding of the worldviews of culturally-different participants and stakeholders in the evaluation, and the use of appropriate evaluation strategies and skills in working with culturally different groups. Diversity may be in terms of race, ethnicity, gender, religion, socio-economics, or other factors pertinent to the evaluation context.

3. Evaluators should practice within the limits of their professional training and competence, and should decline to conduct evaluations that fall substantially outside those limits. When declining the commission or request is not feasible or appropriate, evaluators should make clear any significant limitations on the evaluation that might result. Evaluators should make every effort to gain the competence directly or through the assistance of others who possess the required expertise.

4. Evaluators should continually seek to maintain and improve their competencies, in order to provide the highest level of performance in their evaluations. This continuing professional development might include formal coursework and workshops, self-study, evaluations of one's own practice, and working with other evaluators to learn from their skills and expertise.

C. **Integrity/Honesty:** Evaluators display honesty and integrity in their own behavior, and attempt to ensure the honesty and integrity of the entire evaluation process.

1. Evaluators should negotiate honestly with clients and relevant stakeholders concerning the costs, tasks to be undertaken, limitations of methodology, scope of results likely to be obtained, and uses of data resulting from a specific evaluation. It is primarily the evaluator's responsibility to initiate discussion and clarification of these matters, not the client's.

2. Before accepting an evaluation assignment, evaluators should disclose any roles or relationships they have that might pose a conflict of interest (or appearance of a conflict) with their role as an evaluator. If they proceed with the evaluation, the conflict(s) should be clearly articulated in reports of the evaluation results.

3. Evaluators should record all changes made in the originally negotiated project plans, and the reasons why the changes were made. If those changes would significantly affect the scope and likely results of the evaluation, the evaluator should inform the client and other important stakeholders in a timely fashion (barring good reason to the contrary, before proceeding with further work) of the changes and their likely impact.

4. Evaluators should be explicit about their own, their client’s, and other stakeholders' interests and values concerning the conduct and outcomes of an evaluation.

5. Evaluators should not misrepresent their procedures, data or findings. Within reasonable limits, they should attempt to prevent or correct misuse of their work by others.

6. If evaluators determine that certain procedures or activities are likely to produce misleading evaluative information or conclusions, they have the responsibility to communicate their concerns and the reasons for them. If discussions with the client do not resolve these concerns, the evaluator should decline to conduct the evaluation. If declining the assignment is unfeasible or inappropriate, the evaluator should consult colleagues or relevant stakeholders about other proper ways to proceed. (Options might include discussions at a higher level, a dissenting cover letter or appendix, or refusal to sign the final document.)

7. Evaluators should disclose all sources of financial support for an evaluation, and the source of the request for the evaluation.

D. **Respect for People**: Evaluators respect the security, dignity and self-worth of respondents, program participants, clients, and other evaluation stakeholders.

1. Evaluators should seek a comprehensive understanding of the important contextual elements of the evaluation. Contextual factors that may influence the results of a study include geographic location, timing, political and social climate, economic conditions, and other relevant activities in progress at the same time.

2. Evaluators should abide by current professional ethics, standards, and regulations regarding risks, harms, and burdens that might befall those participating in the evaluation; regarding informed consent for participation in evaluation; and regarding informing participants and clients about the scope and limits of confidentiality.

3. Because justified negative or critical conclusions from an evaluation must be explicitly stated, evaluations sometimes produce results that harm client or stakeholder interests. Under this circumstance, evaluators should seek to maximize the benefits and reduce any unnecessary harms that might occur, provided this will not compromise the integrity of the evaluation findings. Evaluators should carefully judge when the benefits from doing the evaluation or in performing certain evaluation procedures should be foregone because of the risks or harms. To the extent possible, these issues should be anticipated during the negotiation of the evaluation.

4. Knowing that evaluations may negatively affect the interests of some stakeholders, evaluators should conduct the evaluation and communicate its results in a way that clearly respects the stakeholder's dignity and self-worth.

5. Where feasible, evaluators should attempt to foster social equity in evaluation, so that those who give to the evaluation may benefit in return. For example, evaluators should seek to ensure that those who bear the burdens of contributing data and incurring any risks do so willingly, and that they have full knowledge of and opportunity to obtain any benefits of the evaluation. Program participants should be informed that their eligibility to receive services does not hinge on their participation in the evaluation.

6. Evaluators have the responsibility to understand and respect differences among participants, such as differences in their culture, religion, gender, disability, age, sexual orientation and ethnicity, and to account for potential implications of these differences when planning, conducting, analyzing, and reporting evaluations.

E. **Responsibilities for General and Public Welfare:** Evaluators articulate and take into account the diversity of general and public interests and values that may be related to the evaluation.

1. When planning and reporting evaluations, evaluators should include relevant perspectives and interests of the full range of stakeholders.

2. Evaluators should consider not only the immediate operations and outcomes of whatever is being evaluated, but also its broad assumptions, implications and potential side effects.

3. Freedom of information is essential in a democracy. Evaluators should allow all relevant stakeholders access to evaluative information in forms that respect people and honor promises of confidentiality. Evaluators should actively disseminate information to stakeholders as resources allow. Communications that are tailored to a given stakeholder should include all results that may bear on interests of that stakeholder and refer to any other tailored communications to other stakeholders. In all cases, evaluators should strive to present results clearly and simply so that clients and other stakeholders can easily understand the evaluation process and results.

4. Evaluators should maintain a balance between client needs and other needs. Evaluators necessarily have a special relationship with the client who funds or requests the evaluation. By virtue of that relationship, evaluators must strive to meet legitimate client needs whenever it is feasible and appropriate to do so. However, that relationship can also place evaluators in difficult dilemmas when client interests conflict with other interests, or when client interests conflict with the obligation of evaluators for systematic inquiry, competence, integrity, and respect for people. In these cases, evaluators should explicitly identify and discuss the conflicts with the client and relevant stakeholders, resolve them when possible, determine whether continued work on the evaluation is advisable if the conflicts cannot be resolved, and make clear any significant limitations on the evaluation that might result if the conflict is not resolved.

5. Evaluators have obligations that encompass the public interest and good. These obligations are especially important when evaluators are supported by publicly-generated funds; but clear threats to the public good should never be ignored in any evaluation. Because the public interest and good are rarely the same as the interests of any particular group (including those of the client or funder), evaluators will usually have to go beyond analysis of particular stakeholder interests and consider the welfare of society as a whole.