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

---

[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