EVALUATION OF NSF ATE PROGRAM RESEARCH & DEVELOPMENT

In the context of grant-funded programs in science, technology, engineering, and math (STEM) education, confusion has often arisen regarding the definitions and purposes of, and distinctions between, *research* and *evaluation* (e.g., in the National Science Foundation's [NSF's] Advanced Technological Education [ATE] program; Ritchie, 2008). However, such complications may be addressed by applying a research and development paradigm to such projects, as is established by the *Common Guidelines for Education Research and Development* (Institute of Education Sciences [IES], U.S. Department of Education [ED] & the NSF, 2013). This orientation reframes "research" as *research and development* (or R&D), a process whereby an innovation is iteratively improved from an idea grounded in basic research, to resources, materials, and/or technologies with evidence of effectiveness sufficient to warrant wide adoption. Note that the word *innovation* is used here purposefully as a generic term for the treatment, intervention, or solution being developed.

Clarifying research in this way allows "evaluation" to be elaborated more specifically as *program evaluation*—assessment of the *implementation* of activities (including both research and development in this context) and the *impacts* of those activities. The resulting distinction between evaluation and research has substantial promise for clarifying functions, roles, and responsibilities and ultimately improving the success of grant-funded STEM education projects.

The Common Guidelines define six types of research project, any of which should be the subject of a program evaluation if supported by a grant award, and typically executed by a firm or individual external to the R&D activities. All types of research in this model are actually intended to serve two purposes, at least to some degree: not only informing development of the innovation (again, R&D), but also generating broader knowledge or understandings to contribute to a field of practice in education. Program evaluation of such projects should examine success in terms of both of these purposes, assessing implementation and impacts across both research and development activities.

Design and Development Research (Type #3 in the Common Guidelines typology) is of particular interest in this context, given that expectations for the most commonly awarded grant-funded STEM education projects typically stipulate design and iterative development of an innovation, grounded in some existing theory and evidence, and also tested in ways useful to informing ongoing improvement.

Popular approaches to Design and Development Research (or DDR) include *Design-Based Research* (DBR; The Design-Based Research Collective, 2003; Sandoval & Bell, 2004) and *Design-Based Implementation Research* (DBIR; Penuel, Fishman, Cheng, & Sabelli, 2011).

Developmental Evaluation (Patton, 1994; 2011) is a related approach that includes less of an imperative to generate broader understandings for a field, instead focusing more narrowly on informing development. It should be noted, however, that this nomenclature is inconsistent with the distinctions between R&D and program evaluation

central to the orientation described in this document, so its use may reintroduce the complications this approach attempts to avoid.

External evaluation of projects applying an R&D approach can be challenging. Typical methods for evaluating DDR include the convening of expert panels, serving a function much like that provided by a thesis or dissertation committee. This model counts on panelists' expertise rather than a process for systematic inquiry, so may be open to greater criticisms regarding consistency and quality. Other monitoring-focused approaches may document *if or when research is done* (or is done on time), but neglect the *quality* of the undertaking and products of R&D projects.

A different, more structured, approach might define questions (both broad and more detailed) to organize collection of evaluation data, structured by a conceptual framework developed or adopted for the purpose. Hezel Associates, LLC, has developed two such frameworks, both tailored specifically to the potential challenges of evaluating education research and development efforts. Both are targeted at Design and Development Research (Type #3) but may be adapted to other types of R&D described by the Common Guidelines.

Hezel Associates Framework for Evaluating Design and Development Research

Optimized for NSF research projects under the Common Guidelines for Education Research and Development (IES, ED & NSF, 2013)

1. Components of the Research (*Process/Implementation Evaluation*)

    1.1. Development of the innovation

        1.1.1. Specification of a theory of action or logic model

        1.1.2. Alignment of the innovation with the theory of action

        1.1.3. Identification/clarification of, and alignment with, end users

    1.2. Creation of measures to assess the implementation of the innovation

    1.3. Collection of data on the feasibility of implementing the innovation in typical delivery settings by intended users

    1.4. Conducting of pilot studies to examine the promise of the innovation to generate the intended outcomes

2. Research Plan and Execution (*Evaluation of Research Quality*)

    2.1. Method for developing the innovation to the desired point of maturity (an iterative development process)

    2.2. Method for collecting evidence on the feasibility that end users can implement the innovation in authentic settings (evidence of feasibility of implementation)

    2.3. Method for obtaining pilot data on the promise of the innovation for achieving the expected outcomes (pilot study)

3. Project Outputs/Products (from Common Guidelines *Project Outcomes*)

    3.1. Fully developed version of the proposed design-research (including all materials necessary for its implementation)

    3.2. A well-specified theory of action, including evidence supporting or refuting key assumptions of the innovation's original theoretical bases

    3.3. Descriptions of the major design iterations and the resulting evidence to support or question key assumptions about the theory of action

    3.4. Description and empirical evidence of the adjustments to the theory of action and innovation design that resulted from design testing

    3.5. Measures for assessing the innovation in an authentic delivery setting

        3.5.1. Evidence of the technical quality of measures

        3.5.2. Data demonstrating quality and quantities of implementation of the innovation

        3.5.3. Data demonstrating the innovation's level of success in such implementation

    3.6. Pilot data on the innovation's promise for generating the intended beneficial learner outcomes

4. Broader Impacts (NSF merit criterion adapted from *Policy and/or Practical Significance*)

    4.1. Specification of the practical problem the innovation intends to address

    4.2. The importance of that problem, as addressed by the innovation

    4.3. Distinction of how the innovation and its strategies differ from existing practice

    4.4. Explanation of why the innovation has the potential to improve learning or education outcomes, or increase efficiencies in the education system or institutional setting beyond what current practice provides

5. Intellectual Merit (NSF merit criterion adapted from *Theoretical and Empirical Basis*)

    5.1. Description of the contribution to the collective knowledge base

    5.2. Summary of success of the innovation relative to its proposed theoretical/empirical rationale

    5.3. Summary of changes to the theory of action and implications for future research

    5.4. Clarification of relationships (hypotheses) among components of the logic model or theory of action, to inform future research

    5.5. Clarity in the description of those relationships, both theoretically and operationally

<u>Evaluation Questions Addressed by the Hezel Associates Framework</u>

Elaborated by the subheadings above

1. To what extent were components of the design and development research design consistent with expectations framed by the Common Guidelines for Education Research and Development?

2. With what quality and timeliness were research and development activities implemented?

3. What was the quality of outputs or products resulting from the research and development effort?

4. What is the potential for broader impacts to be realized by the innovation being developed and studied?

5. What is the intellectual merit of the research and development effort, in terms of its contributions to understandings about learning?

Evidence Framework for Design Based Implementation Research (DBIR)

Adapted from Means and Harris (2013)

A second option is an adaption of material introduced as an *evidence framework for DBIR*, posited by Barbara Means and Christopher Harris (2013) at an American Educational Research Association national conference.

1. Application of the DBIR approach defined

   1.1. Jointly negotiates a research agenda and questions with practitioners and developers partnering in the research

   1.2. Establishes appropriately flexible hypotheses regarding the theory of action of the innovation being studied

   1.3. Accommodates the need for the innovation to undergo modification during the study

   1.4. Treats the innovation being studied as a set of practices adaptable to local circumstances

   1.5. Seeks to understand implementation as an object of research in and of itself, beyond simply documenting "fidelity"

   1.6. Anticipates inconsistent outcomes across settings (with implications for research data collection)

   1.7. Establishes and applies strategies for identifying unanticipated or unintended consequences

   1.8. Collects data consistent with DBIR design considerations and the developing theory of action of the innovation being studied

   1.9. Manages and analyzes collected data with appropriate quality and rigor, and on anticipated timelines

2. Engagement of co-developers to iteratively refine the innovation being studied

   2.1. Engages in multiple cycles of design, implementation, and refinement

   2.2. Examines the problems being addressed by the innovation

   2.3. Co-develops solutions to those problems

   2.4. Engages educators, students, and others in DBIR methods as they are applied

   2.5. Involves all groups to effectively reflect on and negotiate (a) refinements to the innovation and (b) design principles that may generalize more broadly

3. Generation of research findings that build knowledge regarding STEM learning

   3.1. Understands local context, actions, and outcomes across the learning research sites

   3.2. Develops implementation theory to explain differences in outcomes across research sites

   3.3. Uses those understandings and theories to make appropriate but informative claims for generality that are useful in principle

   3.4. Generates findings useful to inform policy and make decisions that affect education

   3.5. Establishes the range of desirable and acceptable variations in how the innovation is implemented

   3.6. Contributes to understandings of conditions for implementation effectiveness

   3.7. Considers theories of institutional change and organizational learning

   3.8. Contributes to understandings of how to bring the innovation to scale without diluting its effectiveness

   3.9. Disseminates findings to external audiences through the timely development and distribution of quality publications and other resources

4. Increasing the capacity of both the researcher and practitioners participating in the project

   4.1. Researchers become more adept at targeting important-but-challenging issues for education systems

   4.2. Researchers better understand how to conduct rigorous research within the constraints of practicing education systems

   4.3. Collaborating STEM learning program staff become more interested in and inclined toward using data to inform implementation, development, and broader understandings of their work

   4.4. Participating STEM teaching practitioners become more adept at collecting data about both their implementation practices and the outcomes for their programs and specific innovations being applied

Evaluation Questions Addressed by the Evidence Framework for DBIR

Elaborated by subheadings above

1. How, and how effectively, do the Principal Investigator (PI) and R&D team apply the DBIR approach defined for the research project?

2. How, and how effectively, do the PI and researchers engage co-developers to iteratively refine the innovation being studied?

3. How does the DBIR study generate research findings that build knowledge regarding STEM learning?

4. How does the DBIR effort increase the capacity of both the researchers and practitioners participating in the project?

While not strictly aligned to them, the above questions are also consistent with *core principles of DBIR* (Penuel et al., 2011) and recommended *processes for the effective application of design-based research* (Amiel & Reeves, 2008).

By answering these questions, the external evaluation team expects to complement the PI's efforts to reflectively address implications greater than specific questions about the pathways being developed, and to generate broader useful knowledge about uses of digital media resources in STEM learning settings and about the DBIR processes applied in this project (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; van den Akker, Bannan, Kelly, Nieveen, & Plomp, 2007).

Data collection and analysis for such an evaluation will likely apply a *mixed methods approach* (Creswell & Plano Clark, 2011; Greene, 2007), examining information from (a) review of project documents and artifacts (perhaps using rubrics or checklists); (b) expert review (e.g., external advisory panels convened to provide substantive oversight); (c) interviews with the PI and key contributors to the research and development effort; (d) on-site observations of activities at planning meetings; (e) critical review of products of the research study (e.g., descriptive statistics, internal summaries, and reports to be disseminated to external audiences); and (f) peer review, as might result from publication of findings. The choice of data sources will likely be influenced by the evaluation question being answered, per the following table:

Data Needs Table

	Evaluation Questions		Evaluation Data Collection
	Hezel Associates Framework for Evaluating DDR	*Evidence Framework for DBIR*	*Data sources likely to be useful to assess implementation or impact*
Implementation	1. Consistency of research components with the Common Guidelines?	1. Application of the DBIR approach?	• Document and artifact review (e.g., rubric or checklist) • Expert review (e.g., advisory panel) • PI and R&D team interviews
Implementation	2. Implementation quality and timeliness?	2. Engagement of co-developers to iteratively refine the innovation?	• Observation (e.g., working sessions) • Document and artifact review (e.g., planning documents) • R&D team interviews
Implementation	3. Quality of products of R&D effort?		• Expert review of products • Document and artifact review
Impact	4. Potential for impacts of the innovation (broader impacts)?		• Expert review of research findings • Peer review (e.g., publication)
Impact	5. Contributions to understandings about learning (intellectual merit)?	3. Generation of research findings that build knowledge?	• Expert review of research • Peer review • Document and artifact review
Impact		4. Increased capacity of researchers and practitioners?	• PI and R&D team interviews • Observation (e.g., planning meetings) • Document and artifact review

Finally, if a program focuses on delivery of activities without being research (addressing neither of the purposes described in this document), it may be appropriate to simply evaluate the implementation and impact of delivery. This is the case in the NSF ATE solicitation (#14-577), as some tracks stipulate activities and evaluation in ways defining R&D projects, but others do not. The former are designated by research type in the following table; the latter are noted as calling for implementation-impact program evaluation.

Research Alignment for the NSF ATE Program

NSF Program Guidance to Common Guidelines Types

ATE	Program Track	*From the Request for Proposals (RFP): All projects and centers carry out evaluative activities.*	Research Type and Considerations
ATE Projects	Program Development and Improvement	*... resulting program should constitute a model that could be disseminated broadly. Evaluative activities should provide evidence on the extent to which the project goals and objectives are realized.*	R&D; type is dependent on maturity of the model
	Curriculum and Educational Materials Development	*... curriculum and materials development with the intent of nationally disseminating the developed products. ... products will be developed... validated by experts... field tested in diverse locations, and validated in terms of their effectiveness...*	Type #3: Design and Development Type #4: Efficacy Type #5: Effectiveness
	Professional Development for Educators	*Evaluation should demonstrate use in the classrooms and sustainable changes in practice of participating faculty and teachers. Changes in student learning outcomes as well as students' perceptions of technical careers should be measured.*	Program Evaluation; Implementation-Impact
	Leadership Capacity Building	*Not Applicable (NA)*	Program Evaluation; Implementation-Impact
	Teacher Preparation	*... evaluation plan must measure the effectiveness of efforts to recruit prospective K-12 teachers, transfer those students into four-year teacher preparation programs, enhance their understanding of advanced technologies used in the workplace, and enhance their ability to improve the technological literacy of their students.*	Program Evaluation; Implementation-Impact
	Business and Entrepreneurial Skills Development	*NA*	Program Evaluation; Implementation-Impact
	Small Grants for Institutions New to the ATE Program	*... some of the funded projects in this category will serve as a prototype or pilot for an idea that may be expanded in a future proposal for an ATE project.*	R&D; Type #3: Design and Development
	Conferences and Workshops	*... conferences and workshops will be outcome based, and that the final report should contain a statement of the impacts of the event 12-18 months after completion...*	Program Evaluation; Implementation-Impact
	ATE Coordination Networks	*NA*	Program Evaluation; Implementation-Impact
ATE Centers	National Centers Regional Centers Support Centers	*Evaluation of the center's materials and services and their impact on student learning, faculty, and the center's impact on employers and on the institutions that manage the center including longitudinal studies that examine students' performance in the workplace and measure employers' satisfaction with graduates.*	Program Evaluation; Implementation-Impact

	Planning	*Conducting Design Research*	Type #2: Early-Stage or Exploratory Type #3: Design and Development (no pilot)
Targeted Research on Technician Education		*Pilot study*	Type #3: Design and Development (w/pilot)
	Exploratory Research and Development	*… may be built on results from a pilot study or design research study.*	Type #4: Efficacy Type #5: Effectiveness
	Full Scale Research and Development	*… expected to include research on and implementation with other types of participants, at other locations, under different conditions to test development efforts or innovations.*	Type #6: Scale-up

References

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Educational Technology & Society, 11*(4), 29-40.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

The Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher, 32*(1), 5-8.

Greene, J. C. (2007). *Mixed methods in social inquiry.* San Francisco, CA: Jossey-Bass Publishers.

Institute of Education Sciences, U.S. Department of Education, & the National Science Foundation. (2013). *Common guidelines for education research and development.* Retrieved from http://www.nsf.gov/pubs/2013/nsf13126/nsf13126.pdf

Means, B., & Harris, C. J. (2013). *The nature of evidence in design-based implementation research.* Poster presented at the annual meeting of the American Educational Research Association, San Francisco, CA. Retrieved from http://www-personal.umich.edu/~fishman/DBIR-AERA-2013/downloads/files/Means_Harris%20DBIR%20AERA.pdf

Patton, M. Q. (1994). Developmental evaluation. *Evaluation Practice, 15*(3), 311-319.

Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use.* New York, NY: The Guilford Press.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher, 40*(7), 331–337.

Ritchie, L. A. (2008). *Facilitating an understanding of Advanced Technological Education targeted research needs.* Retrieved from http://www.colorado.edu/ibs/decaproject/pubs/FINALATETargetedResearchWhitePaper.pdf

Sandoval, W. A., & Bell, P. (2004). Design-based research methods for studying learning in context: Introduction. *Educational Psychologist, 39*(4), 199-201.

van den Akker, J., Bannan, B., Kelly, A. E., Nieveen, N., & Plomp, T. (2007). *An introduction to educational design research.* Proceedings from the seminar conducted at the East China Normal University, Shanghai (PR China). Retrieved from http://www.slo.nl/downloads/2009/Introduction_20to_20education_20design_20research.pdf/

Citation: (Knestis, 2014)

Knestis, K. (2014). *Evaluation of NSF ATE program research and development.* Syracuse, NY: Hezel Associates.