Common Guidelines for Education Research and Development

A Report from the Institute of Education Sciences,
U.S. Department of Education

and the National Science Foundation

August 2013

Contents

Tables

Preface

In January 2011, a Joint Committee of representatives from the U.S. Department of Education (ED) and the U.S. National Science Foundation (NSF) began work to establish cross-agency guidelines for improving the quality, coherence, and pace of knowledge development in science, technology, engineering and mathematics (STEM) education. The committee formed to enhance the efficiency and effectiveness of both agencies' STEM education research and development programs in response to recommendations from the Office of Science and Technology Policy (OSTP) and guidance from the Office of Management and Budget (OMB) (Zients, 2012). Although the starting place for the committee was research in STEM, ED quickly realized the broader applicability of the guidelines to other content areas in which it funds research and development.

Education research and development programs at NSF are distributed throughout its science and engineering directorates but are located primarily in its Directorate for Education and Human Resources (EHR). EHR's purview includes K-12 education, postsecondary education, and after-school and informal learning environments, as well as the study of science and engineering innovations that emerge from other directorates. ED's research, development, and evaluation programs are located primarily in the Institute of Education Sciences (IES) but also are represented in the Policy and Program Studies Service (PPSS), the Office of Innovation and Improvement (OII), and the National Institute on Disability and Rehabilitation Research (NIDRR).

The Joint Committee examined whether the agencies' expectations for the research studies they fund could be characterized in such a way as to provide cross-agency guidance for program officers, prospective grantees, and peer reviewers. A first task was to define the types of ED- and NSF-funded research that relate to the development and testing of interventions and strategies designed to increase learning. Types of research range from early knowledge-generating projects to studies of full-scale implementation of programs, policies, or practices. Importantly, the committee sought to create a common vocabulary to describe the critical features of these study types to improve communication within and across the agencies and in the broader education research community.

Second, the Joint Committee specified how the types of research relate to one another and described the theoretical and empirical basis needed to justify each research type. The committee emphasizes the importance of proposed studies building on and referencing an evidence base and, in turn, contributing to the accumulation of empirical evidence and development of theoretical models. Throughout its work, the Joint Committee generally adhered to the guiding principles identified in *Scientific Research in Education* (National Research Council, 2002), which call for research that:

- poses significant questions that can be investigated empirically;
- links empirical research to relevant theory;
- uses research designs and methods that permit direct investigation of the question;
- is guided by a coherent and explicit chain of reasoning;
- replicates and generalizes across studies; and
- attends to contextual factors.

Through this document, the Joint Committee seeks to provide a broad framework that clarifies research types and provides basic guidance about the purpose, justification, design features, and expected outcomes from various research types. In that spirit, the Joint Committee intends this to be a "living document" that may be adapted by agencies or divisions within agencies in response to

their needs and opportunities. Over time, the framework may be elaborated or rearranged according to agency focus and assessments of the needs of education researchers and practitioners.

The draft guidelines were distributed throughout ED and NSF for review and comment. NSF held several sessions for agency staff to provide comments and feedback. The agencies jointly sought feedback from the research community at the 2013 annual meetings of the American Educational Research Association, where representatives from ED and NSF presented the guidelines and held small discussion groups. ED and NSF representatives also presented the guidelines at a meeting of Federal evaluators hosted by the Office of Management and Budget (OMB). Finally, NSF leadership reviewed and commented on the document, and detailed reviews of the document by education research experts were obtained through the Institute of Education Sciences' Standards and Review Office.

Members of the Joint Committee

Janice Earle, Co-Chair
Directorate for Education and Human Resources, National Science Foundation

Rebecca Maynard, Co-Chair (2011-2012)
Ruth Curran Neild, Co-Chair (2012-2013)
National Center for Evaluation and Regional Assistance, Institute of Education Sciences

Ex-Officio
John Q. Easton,
Director, Institute of Education Sciences

Joan Ferrini-Mundy
Assistant Director for Education and Human Resources, National Science Foundation

Elizabeth Albro, National Center for Education Research, Institute of Education Sciences

Jinfa Cai, Directorate for Education and Human Resources, National Science Foundation (2011-2012)

Karen Cator, Office of Education Technology, Department of Education (2011-2012)

Gavin Fulmer, Directorate for Education and Human Resources, National Science Foundation (2011-2012)

Edith Gummer, Directorate for Education and Human Resources, National Science Foundation

Jim Hamos, Directorate for Education and Human Resources, National Science Foundation

Michael Lach, Office of the Secretary, Department of Education

Joy Lesnick, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences

Lynn Okagaki, National Center for Education Research, Institute of Education Sciences (2011-2012)

Janet Kolodner, Directorate for Education and Human Resources, National Science Foundation

Jefferson Pestronk, Office of Education Innovation and Improvement, Department of Education (2011-2012)

Anne Ricciuti, Office of the Director, Institute of Education Sciences

Tracy Rimdzius, National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences

Allen Ruby, National Center for Education Research, Institute of Education Sciences

Deborah Speece, National Center for Special Education Research, Institute of Education Sciences

Susan Winter, Social and Behavioral and Economic Sciences Directorate, National Science Foundation (2012)

Introduction

> *At its core, scientific inquiry is the same in all fields. Scientific research, whether in education, physics, anthropology, molecular biology, or economics, is a continual process of rigorous reasoning supported by a dynamic interplay among methods, theories, and findings. It builds understanding in the form of models or theories that can be tested.*
>
> ***Scientific Research in Education***
> National Research Council, 2002

Each year, the National Science Foundation (NSF) and the U.S. Department of Education (ED) make substantial investments in education research and development. Through these efforts, the agencies seek to improve opportunities to learn science, mathematics, engineering, and technology (STEM) and to increase student achievement, engagement and persistence in those areas. ED also supports research and evaluation in a range of areas other than STEM.

Though complementary, the agencies' focus areas in education research differ in ways that correspond to their respective roles in government and society. NSF, which is charged with increasing the quality and amount of science and engineering research in a variety of contexts, has emphasized basic research on STEM learning, cognition, and development of instructional approaches, technologies, and materials in both formal and informal settings. In contrast, ED concentrates its investments on developing and testing the effectiveness of well-defined curricula, programs, and practices that could be implemented by schools. The complementary missions of the agencies, along with the continuing urgency of improving American students' STEM knowledge and skills, form the backdrop for the evidence guidelines and study types described in this document.

This document describes NSF and ED's shared understandings of the roles of various types or "genres" of research in generating evidence about strategies and interventions for increasing student learning. These research types range from studies that generate the most fundamental understandings related to education and learning (for example, about brain activity), to research that examines associations between variables, iteratively designs and tests components of a strategy or intervention, or is designed to assess impact of a fully-developed intervention on an education-related outcome. More specifically, the document describes the agencies' expectations for the purpose of each type of research, the empirical and/or theoretical justifications for different types of studies, types of project outcomes, and quality of evidence.

Fundamentally, these shared, cross-agency expectations are intended to (1) help organize and guide NSF's and ED's respective decisions about investments in education research and (2) clarify for potential grantees and peer reviewers the justifications for and evidence expected from each type of study, as well as relevant aspects of research design that would contribute to high-quality evidence. The primary audiences for this document are agency personnel, scientific investigators who seek funding from these agencies for education research projects, and those who serve as peer reviewers of proposals for scientific research.

By delineating common expectations for study characteristics, it is hoped that each agency will be better able to build on the investments of the other and to see its own investments reap greater

return in improved and tested education practices and policy. And by clarifying the products that should result from different types of studies, the agencies hope to speed the pace of research and development in education—including obtaining meaningful findings and actionable results—through a more systematic development of knowledge (Shonkoff, 2012).

For example, a project that involves design and development of an intervention or strategy should, at its conclusion, have generated a theory of action, a set of intervention components, and preliminary evidence regarding promise for improving education outcomes. In combination, these products from design and development research would make the case that an efficacy trial of a strategy or intervention is warranted, assuming positive and substantively important impacts (see Table 3 for a full set of project outcomes). Without attention to each of these project outcomes, which serve as justification for potentially more-costly and wider-scale testing, the full evidentiary potential of an investment in design and development may not be realized. Likewise, a well-conducted study of impacts should include hypothesis-generating exploratory analyses that can inform additional work. Research on implementation, adaptation, and adoption is an important part of all research endeavors.

Ultimately, these expectations should advance knowledge by asking neither too little nor too much of proposed studies. Too little can be asked of a study when it is not adequately justified or carefully designed to generate good evidence. Too much can be asked when the role of a particular kind of study in evidence generation is unclear. For example, a project about design and development of an intervention should not be required to provide strong evidence of effectiveness among a wide range of populations. If an opportunity for such integration of research purposes occurs, it may be advisable to pursue; however, it also is acceptable for a design and development project to stop short of conducting an efficacy study.

A Cross-Agency Project

This document resulted from collaborations between representatives from the National Science Foundation (NSF) and the U.S. Department of Education (ED) to identify the spectrum of study types that contribute to development and testing of interventions and strategies, and to specify expectations for the contributions of each type of study. This collaboration is but one example of increasing use of evidence government wide to support decision making about investments in programs and research. Although NSF and ED focused on increasing knowledge related to learning in STEM, the general approach described in this document applies to knowledge generation in other areas of education research.

Types of Research

Most simply, the six types of research described in this document form a "pipeline" of evidence that begins with basic and exploratory research, moves to design and development of interventions or strategies, and, for interventions or strategies with initial promise, results in examination of the effectiveness for improving learning or another related education outcome. However, as we describe later in this document, the reality of scientific investigation is more complicated, less orderly, and less linear than such a "pipeline" suggests. In addition, these research types do not represent the entire panoply of useful investigations in education, nor does this document describe the full range of purposes for which a given type of research is useful.

Below, we provide a basic description of the purpose of each of the six types of research. The research types are described in more detail in Tables 1-4.

Foundational Research and Early-Stage or Exploratory Research contributes to *core knowledge* in education. *Core knowledge* includes basic understandings of teaching and learning, such as cognition; components and processes involved in learning and instruction; the operation of education systems; and models of systems and processes.

- ***Research Type #1: Foundational Research*** provides the fundamental knowledge that may contribute to improved learning and other relevant education outcomes. Studies of this type seek to test, develop, or refine theories of teaching or learning and may develop innovations in methodologies and/or technologies that will influence and inform research and development in different contexts.
- ***Research Type #2: Early-Stage or Exploratory Research*** examines relationships among important constructs in education and learning to establish logical connections that may form the basis for future interventions or strategies to improve education outcomes. These connections are usually correlational rather than causal.

Design and Development Research (Research Type #3) develops solutions to achieve a goal related to education or learning, such as improving student engagement or mastery of a set of skills. Research projects of this type draw on existing theory and evidence to design and iteratively develop interventions or strategies, including testing individual components to provide feedback in the development process. These projects may include pilot tests of fully developed interventions to determine whether they achieve their intended outcomes under various conditions. Results from these studies could lead to additional work to better understand the foundational theory behind the results or could indicate that the intervention or strategy is sufficiently promising to warrant more-advanced testing.

Efficacy, Effectiveness, and Scale-up Research contributes to evidence of impact, generating reliable estimates of the ability of a fully-developed intervention or strategy to achieve its intended outcomes. The three types of *Impact Research* share many similarities of approach, including designs that eliminate or reduce bias arising from self-selection into treatment and control conditions, clearly specified outcome measures, adequate statistical power to detect effects, and data on implementation of the intervention or strategy and the counterfactual condition. However, these studies vary with regard to the conditions under which the intervention is implemented and the populations to which the findings generalize. Specifically,

- ***Research Type #4: Efficacy Research*** allows for testing of a strategy or intervention under "ideal" circumstances, including with a higher level of support or developer involvement than would be the case under normal circumstances. Efficacy Research studies may choose to limit the investigation to a single population of interest.
- ***Research Type #5: Effectiveness Research*** examines effectiveness of a strategy or intervention under circumstances that would typically prevail in the target context. The importance of "typical" circumstances means that there should not be more substantial developer support than in normal implementation, and there should not be substantial developer involvement in the evaluation of the strategy or intervention.
- ***Research Type #6: Scale-up Research*** examines effectiveness in a wide range of populations, contexts, and circumstances, without substantial developer involvement in implementation or evaluation. As with Effectiveness Research, Scale-up Research should be

carried out with no more developer involvement than what would be expected under typical implementation.

For each of these research types, the Joint Committee has characterized

- the **purpose**, or how the type of research contributes to the evidence base (Tables 1 and 2);
- the **theoretical and empirical justifications** required for conducting this type of research (Table 3);
- **expectations for research design and expected products** of the research, such as exploratory analysis, impact estimates, or a well-elaborated theory of action (Table 4); and
- **expectations for review** of the products from each type of research (Table 5).

Knowledge Generation and the Complex Connections among Research Types

Although the six study types follow a logical sequence of development of basic knowledge, design, and testing, the Joint Committee emphasizes the reality of building knowledge is considerably more complex. Specifically, it assumes the following:

Knowledge development is not linear. The current of understanding does not flow only in one direction (that is, from basic research to studies of effectiveness). Rather, research generates important feedback loops, with each type of research potentially contributing to an evidence base that can inform and provide justification for other types of research. For example, just as Foundational Research can contribute to a justification for an Impact Research, so can the findings from Impact Research identify needs for more fundamental exploration.

Investigation can sometimes move directly from development of core knowledge to Scale-up Research. New learning opportunities and technologies—Massive Open Online Courses (MOOCs), for example—make it possible to quickly test learning innovations at scale without prior small-scale testing.

Individual studies may incorporate elements that cut across research types. For example, a Design and Development Research project may incorporate a small-scale study to assess efficacy. Likewise, researchers conducting Efficacy Research may need to engage in design and development cycles, and studies of foundational theories of learning may incorporate both elements.

The Joint Committee makes no assumption about the number of studies that will be conducted to address a given education research problem. Sometimes large numbers of Foundational, Early-Stage or Exploratory, or Design and Development studies may be required to develop a strategy or intervention that is ready for wider-scale examination of impact.

Introduction to Tables

The following tables represent the various types of education research studies that were identified, along with how each type might address the following characteristics:

1. Purpose of Foundational, Early-Stage or Exploratory, and Design and Development Research Studies
2. Purpose of Studies that Assess the Impact of Education Interventions and Strategies
3. Justification Guidelines
4. Guidelines for Evidence to Be Produced by Studies
5. Guidelines for External Feedback Plans

In Appendix B, the same information is presented in a different format. Appendix B contains information organized by type rather than characteristics.

Table 1: Purpose of Foundational, Early-Stage or Exploratory, and Design and Development Research Studies

Foundational, Early-Stage or Exploratory, and Design and Development Studies	
An important genre of education research advances knowledge about fundamental principles of sound education practice. This includes studies to advance foundational knowledge that guides theory development; Early-Stage or Exploratory research to identify evidence of the promise (or lack thereof) of programs, policies or practices; and research that guides the development and early-stage testing of innovative programs, policies and practices to improve education outcomes.	
Foundational Research	The purpose of Foundational Research is to advance the frontiers of education and learning; develop and refine theory and methodology; and provide fundamental knowledge about teaching and/or learning. Foundational Research studies may examine phenomena without establishing an explicit link to education outcomes.
Early-Stage or Exploratory Research	The purpose of Early-Stage or Exploratory Research is to investigate approaches to education problems to establish the basis for design and development of new interventions or strategies, and/or to provide evidence for whether an established intervention or strategy is ready to be tested in an efficacy study. Early-Stage or Exploratory Research should establish initial connections to outcomes of interest. Studies in this genre should support the development of a well-explicated theory of action that can inform the development, modification, or evaluation of an intervention or strategy. They should build on existing research and theory to examine issues such as: (1) Associations between (a) education or learning outcomes and (b) malleable factors (that is, factors that are alterable, such as children's behaviors; technologies; education programs, policies, and practices) and; (2) Factors and conditions that may mediate or moderate the relationship between (a) education or learning outcomes and (b) malleable factors; and (3) Opportunities for new interventions or strategies, and challenges to their adoption, with the goal of informing policy, practice, and future design or development.
Design and Development Research	The purpose of Design and Development Research is to develop new or improved interventions or strategies to achieve well-specified learning goals or objectives, including making refinements on the basis of small-scale testing. Typically this research involves four components: (1) Development of a solution (for example, an instructional approach; design and learning objects, such as museum exhibits or media; or education policy) based on a well-specified theory of action appropriate to a well-defined end user; (2) Creation of measures to

assess the implementation of the solution(s); (3) Collection of data on the feasibility of implementing the solution(s) in typical delivery settings by intended users; and (4) Conducting a pilot study to examine the promise of generating the intended outcomes.

In some cases, funders will expect all four stages to be completed within a single project; in other cases, Design and Development Projects may entail sequential projects.

Table 2: Purpose of Studies that Assess the Impact of Education Interventions and Strategies

Studies of Impact	
The purpose of Impact Studies is to generate reliable estimates of the ability of a *fully developed* intervention or strategy to achieve its intended outcomes. For an impact study to be warranted, the theory of action must be well established and the components of the intervention or strategy well specified. The three types of impact studies—Efficacy, Effectiveness, and Scale-up—differ with regard to the conditions under which the intervention is implemented and the populations to which the findings generalize. In addition, as the research moves from Efficacy to Scale-up, studies should also give greater attention to identifying variation among impacts by subgroup, setting, level of implementation, and other mediators. For all impact studies, descriptive and exploratory analyses should be sufficiently elaborated to determine the extent to which the findings support the underlying theory of action.	
Efficacy Research	The purpose of Efficacy Research is to determine whether an intervention or strategy can improve outcomes under what are sometimes called "ideal" conditions. For example, these conditions may include more implementation support or more highly trained personnel than would be expected under routine practice, or in contexts that include a more homogeneous sample of students, teachers, schools, and/or districts than is typical. Efficacy studies may involve the developer in the implementation of the intervention or strategy; however, the study should include reasonable safeguards for ensuring the objectivity and integrity of the study. Sometimes Efficacy studies are used to replicate previous evaluations of an intervention, but under different conditions (e.g., with a different population or using a variant of the intervention or strategy).
Effectiveness Research	The purpose of Effectiveness Research is to estimate the impacts of an intervention or strategy when implemented under conditions of routine practice. To this end, implementation should be similar to what would occur if a study were not being conducted. An Effectiveness study should be carried out with no more developer involvement than what would be expected under typical implementation.
Scale-up Research	The purpose of Scale-up Research is to estimate the impacts of an intervention or strategy under conditions of routine practice *and* across a broad spectrum of populations and settings. That is, Scale-Up studies should be conducted in settings and with population groups that are sufficiently diverse to broadly generalize findings.

As with Effectiveness Research, Scale-up Research should be carried out with no more developer involvement than what would be expected under typical implementation.

Table 3: Justification Guidelines

Foundational Research	
Policy and/or Practical Significance	The proposed project should address important research problems or questions related to education and learning. Although the project should have clear potential implications for policy and/or practice, a direct relationship to student outcomes is not required.
Theoretical and Empirical Basis	The proposal should outline the theoretical and empirical bases for the project and, if relevant, identify why it is necessary to develop new theory, explore learning constructs, or create more useful conceptual frameworks. The proposal should describe whether and how the project will identify or explore important new constructs in education and learning, extend understanding of current constructs, expand understanding of relationships among the constructs under investigation, and/or extend research methodologies appropriate to advancing the evidence base to support improved policy or practice.
Early-Stage or Exploratory Research	
Policy and/or Practical Significance	The proposal for the project should provide a clear description of the practical education problem or issue that will be the study focus and a compelling rationale for studying the problem. The project should provide a compelling case that the proposed research will generate important knowledge to inform the development, improvement, or evaluation of education programs, policies, or practices.
Theoretical and Empirical Basis	The proposal should detail a strong theoretical and empirical rationale for the project. To the extent possible, an empirical rationale should be included with citations of supporting evidence. When a study of an existing intervention or strategy is proposed, there should be a compelling explanation of why this intervention should be studied through Early-Stage or Exploratory Research rather than through Efficacy Research.
Design and Development Research	
Policy and/or Practical Significance	The proposal for the project should provide a compelling rationale that (1) specifies the practical problem the proposed intervention intends to address; (2) justifies the importance of the problem; (3) describes how the proposed intervention or strategy differs from existing practice; and (4) explains why the proposed project has the potential to improve learning or education outcomes or increase efficiencies in the education system or institutional setting beyond what current practice provides.
Theoretical and Empirical Basis	The proposal should include a strong theoretical and empirical justification for development of the proposed intervention or strategy. If the theoretical basis rests on evidence related to individual features or components, the proposal should provide a compelling rationale for how combining these features or components into a new intervention is expected to achieve intended outcomes. The proposal should include a description of the initial concept for the

	planned investigation, including a well-explicated theory of action or logic model. The concept and logic model should identify key components of the intervention (i.e., the ingredients hypothesized to be critical to achieving the intended results) and should describe their relationships, theoretically, and operationally.
Efficacy Research	
Policy and/or Practical Significance	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under ideal implementation conditions with a well-defined sample, rather than under routine practice conditions with a relevant typical sample or under typical implementation conditions with a broad sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.
Theoretical and Empirical Basis	Efficacy Research should be justified by one or more of the following: (1) empirical evidence of the promise of the intervention from a well-designed and implemented pilot study (e.g., a study conducted as part of a design and development project); (2) empirical evidence from at least one well-designed and implemented Early-Stage or Exploratory Research study supporting all the critical links in the intervention's theory of action; (3) evidence the intervention is widely used even though it has not been adequately evaluated to determine its efficacy; or (4) if the intent is to replicate an evaluation of an intervention with a different population, evidence of favorable impacts from a previous well-designed and implemented efficacy study and justification for studying the intervention with the new target population.
Effectiveness Research	
Policy and/or Practical Significance	The proposal for the project should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under routine practice conditions with a relevant typical sample, rather than under ideal implementation conditions with a

	well-defined sample or under typical implementation conditions with a broad sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.
Theoretical and Empirical Basis	Effectiveness Research should be justified by strong empirical evidence of the efficacy of the intervention, as demonstrated by statistically significant and substantively important estimates of impact, from one study that includes multiple sites or settings, or two studies that each include one site or setting, all of which meet the guidelines for evidence to be produced by Impact Research (Table 4) or evidence that the intervention is widely used even though it has not been adequately evaluated for efficacy.
Scale-up Research	
Policy and/or Practical Significance	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under typical implementation conditions with a broad sample, rather than under ideal implementation conditions with a well-defined sample or under routine practice conditions with a relevant typical sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.
Theoretical and Empirical Basis	Scale-up Research should be justified by compelling evidence of the effectiveness of the intervention, as demonstrated by statistically significant and substantively important impact estimates from one study that includes multiple sites or settings, or two studies that include one site or setting, all of which meet the guidelines for evidence to be produced by Impact Research (Table 4). In addition, there should be no overriding evidence demonstrating a negative impact of the intervention.

Table 4: Guidelines for Evidence to Be Produced by Studies

Foundational Research	
Project Outcomes	The project's potential outcomes should include advances in theory, methodology, and/or understandings of important constructs in education and learning. Foundational Research studies should have appropriate methodological rigor such that, upon completion, the project's findings could serve as the basis for future studies.
Research Plan	The proposal should clearly define (1) key conjectures or hypotheses, research questions, and research objectives that derive from the theoretical and empirical basis of the study; (2) a detailed description of the study design, including but not limited to a description of the population(s) of interest; (3) sampling or selection methods, and the expected sample size; and (4) methods to be used for data analysis. For studies that include hypothesis testing, the proposal should discuss the minimum relevant mean difference or relationship between variables and sample size required to ensure adequate statistical power to detect true differences or relationships of this magnitude or larger. For qualitative studies or study components, proposals should provide a rationale for the sample size and selection plan. For studies that analyze secondary data, the proposal should describe the source and availability of data and the sequence of modeling planned. For studies that collect primary data, the proposal should describe the instruments and protocols that will be developed and used, provide initial evidence from literature to support assumptions that guide the sample design, and describe strategies for ensuring validity and reliability of the outcome measures, and discuss strategies for triangulation of findings. Finally, all proposals should include explicit plans for data management and analysis, including statistical models and/or procedures for analysis of text, video, or observational data, data curating and sharing, and for dissemination of findings.
Early-Stage or Exploratory Research	
Project Outcomes	The project's outcomes should include (1) Empirical evidence regarding (a) the malleable factors' association with education or learning outcomes and/or (b) evidence on whether the factors and conditions moderate and/or mediate the relations between the malleable factors and the learner outcomes; (2) (a) A well-specified conceptual framework that supports a theoretical explanation for the link between the malleable factors and the education or learning outcomes and/or (b) a theoretical explanation for the factors' and conditions' moderation and/or mediation of the relationship between the malleable factors and learner outcomes; and (3) A determination based on the empirical evidence and conceptual framework of whether there is a basis for pursuing a Design and Development project or an Efficacy study or whether further

	Foundational or Early-Stage or Exploratory Research is needed before proceeding to efficacy or effectiveness testing.
Research Plan	The research plan should include the following: (1) a set of hypotheses or research questions that are derived from the theoretical and empirical rationale for the study; (2) a detailed research design that is appropriate for the hypotheses or research questions; (3) a justification for the proposed research context and sample; (4) a description of the data sources if secondary analyses are proposed ; (5) a detailed description of data collection procedures and instruments, as well as evidence of and strategies for ensuring reliability and validity; (6) if applicable, a plan to study the opportunities for interventions (i.e., programs, policies, practices or technologies) to address education and learning challenges; and (7) a detailed description of data analysis procedures and the reporting plan.
Design and Development Research	
Project Outcomes	The project's outcomes should include (1) a fully developed version of the proposed design-research (including all materials necessary for its implementation); (2) a well-specified theory of action, including evidence supporting or refuting key assumptions of the intervention's original theoretical basis ; (3) descriptions of the major design iterations and the resulting evidence to support or question key assumptions about the theory of action; (4) description and empirical evidence of the adjustments to the theory of action and intervention design that resulted from design testing; (5) measures with evidence of technical quality for assessing the implementation of the intervention in an authentic education delivery setting and data demonstrating the project's success in such implementation; and (6) pilot data on the intervention's promise for generating the intended beneficial learner outcomes.
Research Plan	The research plan should describe the (1) method for developing the intervention to the point where it can be used by the intended end-users (iterative development process); (2) method for collecting evidence on the feasibility that end users can implement the intervention in an authentic education or learning setting (evidence of feasibility of implementation); and (3) method for obtaining pilot data on the promise of the intervention for achieving the expected outcomes (pilot study).
Impact Research (Efficacy, Effectiveness, and Scale-up)	
Project Outcomes	Efficacy, Effectiveness, and Scale-up reporting should include detailed descriptions of the study goals, design and implementation, data collection and quality, and analysis and findings, for example, as outlined in the What Works Clearinghouse author reporting guidelines (http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=235). The core findings from these studies should be reliable estimates of the intervention's average impact. In some cases, it also will be possible and desirable to estimate impacts for sample subgroups defined by such characteristics as setting, population group, or cohort.

	Study reports should document implementation of both the intervention and the counterfactual condition in sufficient detail for readers to judge applicability of the study findings. When possible, these factors should be related descriptively to the impact findings. Study reports should discuss implications of the findings for the theory of action and, where warranted, make suggestions for adjusting the theory of action to reflect the study findings. If a favorable impact is found, the project should identify the organizational supports, tools, and procedures that were key features of the intervention implementation. If no evidence of a favorable impact is found, the project should examine possible reasons (e.g., weaknesses in the implementation, evidence that raises questions about particular aspects of the logic model).
Research Plan	The research plan should identify and justify (1) the study design used to estimate causal impact of the intervention on the outcomes of interest; (2) the key outcomes of interest for the impact study and the minimum size impact of the intervention that would have policy or practical relevance; (3) the study setting(s) and target population(s); (4) the sample, including the power it provides for detecting an impact; (5) the data collection plan, including information about procedures and measures, including evidence on and strategies for ensuring reliability and validity, and plans for collecting data on program implementation, comparison group practices, and study context; and (6) the analysis and reporting plan . Efficacy, Effectiveness, and Scale-up research should use study designs that will yield impact estimates with strong causal validity and that, for example, could meet What Works Clearinghouse standards without reservations (http://ies.ed.gov/ncee/wwc/). Generally and when feasible, they should use designs in which the treatment and comparison groups are randomly assigned. For Impact Research (as opposed to Design and Development Research), quasi-experimental designs, such as matched comparison groups or regression discontinuity designs, are acceptable only when there is direct compelling evidence demonstrating the implausibility of common threats to internal validity. These might include selection bias in the case of matched comparison groups, or, in the case of regression discontinuity designs, nonlinearities of treatment impacts over a policy relevant span around the "cut point." Ideally, the study sample size and allocation to condition should be such that the minimum true impact detectable size with 80 percent power and a 95 percent confidence interval is no larger than the minimum relevant size impact for policy or practice. If that is not the case, the proposal should provide a rationale for conducting the study despite its not meeting this standard.

Primary outcome measures should include student outcomes sensitive to the performance change the intervention is intended to bring about (e.g., researcher-developed measures that are aligned with the experiences of the treatment group), student outcomes not strictly aligned with the intervention, and student outcomes of practical interest to educators and policymakers (e.g., test scores, grades, graduation or dropout rates). These outcomes should be pre-specified, have been demonstrated as reliable and valid for the intended purposes, and based on data-collection methods that have been shown to yield reliable data.

The project should measure the strength and qualities of implementation (sometimes referred to as "fidelity of implementation") to address whether the intervention's impact estimates may be linked to how it was implemented.

The project should measure comparison group practices and/or conditions to support a clear characterization of the contrast between the intervention and comparison condition. Proposals for Impact Studies should identify the measures, the validity and reliability of these measures, and how data will be collected on these measures.

The analysis plan should specify analytic models that reflect the sample design and maximize the likelihood of obtaining unbiased, efficient estimates of average impacts and the confidence intervals around those impacts.

The analysis plan should describe additional analyses conducted to explore variability in the intervention's impacts and possible implications for the theory of change. For example, these analyses could include (1) subgroup analyses (expected in Effectiveness and in Scale-up Studies); (2) exploration of co-variation in impact estimates and fidelity of implementation or intervention contrasts; and (3) evidence of possible moderator and mediator effects.

Table 5: Guidelines for External Feedback Plans

For all Research Types	The project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator, and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings.

Conclusion

These *Common Guidelines for Education Research and Development* are expected to provide critical guidance to NSF, ED, and the broader education research-and-development community. The two agencies will draw on the Common Guidelines to communicate the goals of their supported education research projects and to establish uniform expectations for proposals submitted in response to particular program announcements, solicitations, or other funding opportunities. Common Guidelines referenced in program solicitations, for example, may require proposers to address them in their application methodologies. In addition, the Guidelines contain important considerations in planning a project, including building the research team.

Where research plans align Common Guidelines with formal proposal review criteria, agencies must ensure expert review panels are well informed of how the guidelines should be applied when evaluating proposals. For example, these guidelines may give reviewers a tool to assess the quality of the research design, both for individual proposals and across a group of proposal, which will help ensure agencies fund robust research and development efforts.

More generally, it is expected that researchers, developers, and evaluators will need to become familiar with the Common Guidelines to prepare successful proposals, as well as carry out research funded by the two agencies.

These guidelines can help practitioners develop a better understanding of what different stages of education research should address and be expected to produce. This, in turn, can support better-informed decisions based on the level of evidence provided.

Outside of NSF and ED activities, the Common Guidelines are expected to be used by education researchers, materials developers, project and program evaluators, and others. The Common Guidelines also may make the public more aware of the agencies' goals for investments in education research and development to achieve immediate and long-term improvement of education and learning.

Finally, OMB, OSTP, the Government Accountability Office, and other federal entities may elect to use the Common Guidelines as part of executive and Congressional oversight of NSF and ED investments.

References Consulted by the Joint Committee

American Educational Research Association (2006). Standards for reporting on empirical social science research in AERA publications, *Educational Researcher*, Vol. 35, No. 6, pp. 33–40. Retrieved from: http://www.aera.net/Portals/38/docs/12ERv35n6_Standard4Report%20.pdf.

Bryk, A. S. (2009). Support a science of performance improvement. *Phi Delta Kappan,* April 2009, pp. 597-600.

Burkhardt H. and Schoenfeld, A. (2003). Improving educational research: Toward a more useful, more influential and better-funded enterprise*. Education Researcher*, 32 (9), 3-14. DOI: 10.3102/0013189X032009003

Clements, D.H. and Sarama, J. (2009*). Learning and Teaching Early Math: The Learning Trajectories Approach*. NY: Routledge.

Cochrane Collaboration (2007). *The Cochrane Collaboration: The Reliable Source of Evidence in Health Care*. Retrieved from http://www.cochrane.org.

Kelly, A. E. (2004). Design research in education: Yes, but is it methodological? *Journal of the Learning Sciences*, 15 (1), 181-287. Retrieved from JSTOR. Retrieved from http://www.jstor.org/stable/1466935

Institute of Medicine (2008). *Knowing What Works In Health Care: A Roadmap for the Nation*. Washington, DC: The National Academies Press.

Murnane, R. J. and Willett, J. B. (2011). *Methods Matter: Improving Causal Inference in Educational and Social Science*. New York: Oxford University Press.

National Academy of Sciences (2004). *On Evaluating Curricular Effectiveness: Judging the Quality of K-12 Mathematics Evaluations*. Washington, DC: National Academy Press.

National Research Council (2002). *Scientific research in education*. Washington, DC: National Academy Press.

Penuel, W., Fishman, B., Cheng, B. H., and N. Sabelli (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, 40 (7), 331-337. DOI: 10.3102/0013189X11421826

Shonkoff, Jack P. (2012). *The Frontiers of Innovation, Minds Wide Open: An Action Strategy and Status Report on the Frontiers of Innovation in Early Childhood Policy and Practice*, Harvard University Center on the Developing Child, May 2012.

Tseng, V. (2009). *Learning about the Use of Research to Inform Evidence-based Policy and Practice: Early Lessons and Future Directions*. Annual Report. William T. Grant Foundation. New York.

U.S. Department of Education (2011). *Digest of Education Statistics*, 2010. National Center for Education Statistics (NCES 2011-015).

U.S. Department of Education (2009). *The Impacts of Regular Upward Bound on Postsecondary Outcomes Seven to Nine Years After Scheduled High School Graduation*. Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service, U.S. Department of Education, Washington, DC.

U.S. National Science Foundation. (2011). *Mission of the Directorate for Education and Human Resources*. Retrieved from http://www.nsf.gov/ehr/about.jsp

Wilensky, Gail R. (2006). Developing a center for comparative effectiveness information. *Health Affairs,* 25 (6), 572-585. doi: 10.1377/hlthaff.25.w572

Zients, Jeffrey. (2012). *Use of Evidence and Evaluation in the 2012 Budget*, Office of Management and Budget, May 18, 2012. Retrieved from: http://www.whitehouse.gov/sites/default/files/omb/memoranda/2012/m-12-14.pdf

Appendix A: Illustrative Research Projects Funded by the Department of Education or the National Science Foundation

Examples of Research Types

The following examples are drawn from both the NSF and U.S. Department of Education research portfolios. They were chosen by the Joint Committee as examples of the different types of research that are identified in the tables. The Guidelines are intended for proposals. The examples are drawn from projects that have findings and/or publications so that the reader might see how the research was structured and implemented. Through the use of these examples, we hope that readers will gain a richer sense of the study's purpose, research design, justification, and findings.

Research Type: Foundational Research

The role of individual differences in executive function on the child's acquisition of intuitive biology

Deborah Zaitchik, PI

Purpose: The purpose of the project funded by NSF is to explore a novel hypothesis that some intuitive, naïve ways of reasoning may never be truly discarded, but are present and inhibited by other mental actions. If so, then observed conceptual change among older children and healthy adults is due to inhibition of naïve conceptions, rather than due to replacement of these naïve views with more scientifically accurate conceptions. The proposed mechanism for this conceptual change is executive function, which includes for this study working memory, conflict monitoring, and inhibition.

Justification: One of the major pieces of children's development into adults is to set aside *animist biology* (in which something is presumed to live if it moves) and to learn *intuitive biology* (the commonsense theory of living things that most adults hold).

This project is built on prior studies among young children and among adults with cognitive impairments. Among the previous research are findings that: young children engage in animist biological thinking and slowly transition into intuitive biological thinking; adults with Williams syndrome (a form of cognitive disability) have obtained factual knowledge about many aspects of life and living things, but do not understand biological concepts of family/heredity, bodily function, and death; and elderly adults with Alzheimer's disease regress to using animist biological reasoning.

Research Plan: The project will study animist and intuitive biological concepts, and relationships to executive functions, among multiple populations: healthy children, healthy elderly adults, and elderly adults with Alzheimer's disease. [The researchers' previous studies have examined these factors among adults with William syndrome.] All participants will complete a set of tests of executive function, including cognitive control (ability to recognize conceptual conflict between ideas), working memory (how many ideas can the person hold in memory), and inhibition. The participants will also complete interviews and tasks about their biological understandings, including: an animism interview, a death interview, a body interview, and a species transformation

task. The study will recruit 60 children, 20 healthy elderly adults, and 20 elderly adults with Alzheimer's disease.

Findings: Tests of biological reasoning are significantly correlated, indicating that what children are acquiring is a coherent theory rather than an elaborated collection of facts. Individual children's performance on tests of biological reasoningare significantly correlated with their tests of executive function. Interventions that enhance executive function may be particularly effective in helping children undergo the conceptual changes that underlie theory development. The pace of acquisition of vitalist biology, a specific domain of STEM knowledge, is determined in part by domain-general mechanisms such as inhibition, working memory, and setshifting.

Publications:

Zaitchik, D., Iqbal, Y., & Carey, S. (under review) The effect of executive function on biological reasoning in young children: An individual differences study.

Powell, L., Carey, S. (under review). Executive function depletion in children and its impact on theory of mind.

Research Type: Early-Stage or Exploratory Research

Reducing the Racial Achievement Gap in STEM: A Social-Neurobiological Investigation and Values Affirmation (https://arc.uchicago.edu/reese/projects/reducing-racial-achievement-gap-self-affirmation-intervention-approach)

Geoffrey Cohen, PI

Purpose: The purpose of the study funded by NSF is to examine possible physiological mechanisms for identity threat and self-affirmation interventions, and whether an online format is feasible for the same effects.

Justification: There continues to be an achievement gap between African American and white students. Furthermore, prior study has shown that African American students perform less well when their identity as African Americans is emphasized, compared to when they complete the test without any such signals. Prior research and theory on stereotype threat, identity threat, and self-affirmation have examined the ways that being reminded of membership in a stereotypic group can reduce academic performance, and that there are self-affirmative approaches that can reduce or eliminate this effect. What is not known is the exact mechanism by which these interventions affect performance.

Research Plan: The research questions are: What are the mechanisms by which identity threat has a long-term cumulative impact? How can these inform intervention? The project proposes a mechanism through which identity threat activates a physiological response to threat, which will be detected via three stress-related compounds: cortisol, alpha-amylase, and pro-inflammatory cytokines (in this case, TNF-alpha). In a first study, participants (100 undergraduates) will be assigned at random to control or experimental conditions, and will be tested in the stress responses and in academic performance. In a second study (also 100 participants), the experimental conditions for will be implemented via online format for undergraduates in a large-format biology course.

Findings: Values affirmation is effective at improving achievement of minority middle school students, with effects that persist for 2 years. Effect was also generalized beyond suburban African Americans to immigrant, economically disadvantaged Latino Americans.

Publications:

Cohen, G. L., Garcia, J., Purdie-Vaugns, V., Apfel, N., & Brzustoski, P. (2009). Recursive processes in self-affirmation: Intervening to close the minority achievement gap. *Science*, 324, 400-403.

Cohen, G. L., & Garcia, J. (2008). Identity, belonging, and achievement: A model, interventions, implications. *Current Directions in Psychological Science*, 17, 365-369.

Purdie-Vaughns, V., Cohen, G. L., Garcia, J., Sumner, R., Cook, J. C., & Apfel, N. H. (2009). Improving minority academic performance: How a values-affirmation intervention works. *Teachers College Record*, September 23.

Miyake, A., Kost-Smith, L., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. *Science* 330(6008): 1234-1237.

EARLY-STAGE OR EXPLORATORY RESEARCH EXAMPLE 2

An Exploration of Malleable Social and Cognitive Factors Associated with Early Elementary School Students' Mathematics Achievement

(http://ies.ed.gov/funding/grantsearch/details.asp?ID=1201)

Sian Beilock, PI

Purpose: The goal of this project funded by IES is to explore how young students' mathematics anxiety and attitudes relate to their mathematics achievement, as well as the cognitive factors that mediate identified anxiety-achievement relations. In addition, the researchers will explore the relation between teachers' math anxiety and students' math achievement.

Justification: Individuals who have negative attitudes about mathematics are often high in mathematics anxiety. Math anxious individuals also tend to have poor math knowledge and often avoid math courses and math-related career paths. Thus, negative math attitudes can have an adverse effect on students' mathematics and science achievement.

Research Plan: Working with first and second grade teachers and students, the researchers conduct three studies to explore the association between first and second grade students' math anxiety, math attitudes, and math achievement. Of particular focus is the relationship between teachers' math anxiety and their students' math achievement. In all three studies, students' and teachers' reading attitudes, anxiety, and knowledge are examined as a comparison domain. The first study will gather student and teacher data at the beginning and end of the school year to explore how students' math anxiety and attitudes relate to their math achievement. The study also seeks to identify cognitive factors that mediate (e.g., math problem solving strategies) this anxiety-achievement relation. The second study observes teachers' identified with low or high math anxiety as they teach their students in mathematics and reading. The researchers also examine code observed behavior, and explore if those behaviors are related to student achievement. The third study experimentally tests whether exposure to different teacher practices are associated with student math and reading anxiety, stereotypes, and math and reading performance.

Findings: This project is ongoing, and additional results will be forthcoming.

Publications:

Maloney, E., & Beilock, S. L. (2012). Math anxiety: Who has it, why it develops, and how to guard against it. *Trends in Cognitive Science*, *16*, 404-406.

Research Type: Design and Development

Project Words of Oral Reading and Language Development (Project WORLD)

Jorge Gonzalez, PI

(http://ies.ed.gov/funding/grantsearch/details.asp?ID=174)

Purpose: The purpose of this project funded by IES was to develop and test a preschool curriculum designed to bridge the vocabulary and comprehension gap that exists for many disadvantaged children. The curriculum integrates preschool and home learning, narrative and informational text, and word and world knowledge.

Justification: During the preschool years, children develop language skills that influence their long-term ability to read with comprehension. We know that children start school with different levels of language skills, and that those entry-level skills are correlated with long-term performance. However, there are few interventions that specifically target developing comprehension skills among preschool children.

Research Plan: Working closely with teachers, the researchers used a design experiment in Year 1 to systematically develop and formatively refine the English version of the intervention through observations and analysis. A series of iterations were field tested and feedback from teachers and observational data were used to make modifications. The team also developed expressive and receptive measures of vocabulary to measure learning of the targeted vocabulary words, and a set of observational codes to capture student engagement and vocabulary use by teachers and students in the classroom.

In Year 2, the initial English version of WORLD was field tested to assess the feasibility of the full implementation, and the promise of the curriculum on improving student vocabulary. Twenty classrooms were randomly assigned to the WORLD curriculum or a business-as-usual comparison group. The curriculum was delivered by teachers over 12 weeks in 5-day instructional cycles of 20-minutes of shared readings targeting world knowledge, vocabulary, and knowledge of text structure. The team learned that preschoolers with the lower level of vocabulary prior to participation in the intervention benefitted from language opportunities beyond the focused activities of the intervention. Those additional opportunities were positively related to those students' language outcomes. Findings from the initial pilot test indicated positive effects on the researcher-developed vocabulary measures, but not on the standardized vocabulary measures. These findings were then used to revise and improve the curriculum. The primary revision was to increase the intensity of the curriculum, going from 12 to 18 weeks of instruction. Other activities in Year 2 included a second field test comparing the teacher-delivered curriculum to the teacher-delivered plus parent-delivered curriculum, where no benefits of the additional home component were found. Finally, the Spanish versions of the teacher and parent components were developed in Year 2.

In Year 3, a pilot study to test the promise of the revised intervention was completed. The team implemented a randomized field trial of the English teacher- and parent-delivered versions of the intervention was conducted with 21 pre-kindergarten and Head Start teachers and their students randomly assigned at the teacher level to WORLD (n=13) or a business-as-usual comparison group (n=8). A smaller pilot test of Spanish WORLD was also run using pre-post comparisons.

Findings: Findings from the Year 3 pilot study indicated positive effects of the WORLD curriculum on researcher-developed measures and standardized measures in both the randomized field trial of English WORLD and the smaller pilot of Spanish WORLD. In the field trial of the English WORLD curriculum, the expressive vocabulary of students in the intervention group grew approximately 6 months more than the growth seen in the comparison group. A substantial, but smaller amount of growth, was seen in students' receptive vocabulary growth. These at-risk preschool students were close to a year behind where they should be for their age, but this vocabulary gap was narrowed as a result of participation in this intervention. Children's baseline vocabulary scores did not moderate the effect of the curriculum, and the home component did not increase the impact of the curriculum in either the English or Spanish version. Thus, the data from this pilot study have demonstrated promise for reducing the vocabulary gap for at-risk preschool children. Several peer-reviewed publications have been produced from this study.

Publications:

Gonzalez, J. E., Darrensbourg, A., Perez, E., Villareal, V., Kim, M., & Haynes, R. (2011). Exploring the Underlying Factor Structure of the English and Spanish Translation of the "Familia" Family Literacy Inventory: A Cautionary Tale. *Early Childhood Research Quarterly, 26*(4): 475-483.

Gonzalez, J. E., Pollard-Durodola, S., Simmons, D. C., Taylor, A., Davis, M. J., Kim, M., & Simmons, L. (2011). Developing Low-Income Preschoolers' Social Studies and Science Vocabulary Knowledge Through Content-Focused Shared Book Reading. *Journal of Research on Educational Effectiveness*. 4(1): 25-52.

Pollard-Durodola, S., Gonzalez, J. E., Simmons, D., Taylor, A., Davis, M., & Simmons, L. (2011). The Effects of an Intensive Shared Book-Reading Intervention for Preschool Children at Risk for Vocabulary Delay. *Exceptional Children*, 77(2): 161-183.

Simmons, D. C., Pollard-Durodola, S. D., Gonzalez, J. E., Davis, M. & Simmons, L. (2007). The Construction of Shared-Reading Interventions: Principles for Accelerating the Vocabulary Development and Comprehension of Low-Income Households. In. S. B. Neuman (Ed.), *Literacy achievement for young children from poverty* (pp. 187-212). Baltimore, MD: Brooks Publishing.

Crossing the Boundaries of Design and Development and Early Efficacy Research

Deep Think: Thinking Deeply About Biodiversity and Ecology

Nancy Songer, PI

(http://www.soe.umich.edu/research/projects/deepthink_thinking_deeply_about_biodiversity_and_ecology)

Purpose: Learning science consists not just of memorizing a body of facts, but represents a complex interplay of content and scientific practices. The study funded by NSF describes early, intermediate and advanced levels of scientific knowledge development in the conceptual areas of biodiversity and ecology AND in the strategic areas of scientific explanation and data analysis, expanding prior work conducted in the sixth grade to grades four and five. The PIs designed and tested scaffolds to support students in fusing content in ecology with the practice of evidence-based explanations and also developed assessments that address the science conceptual and scientific reasoning areas under study.

A quasi-experimental longitudinal study of the cumulative impact of engaging with these curricular materials and instructional strategies of matched cohorts of students across the three years of the study was also conducted. A parallel study examined the same constructs in the in-service teachers of these students and with preservice teachers in the program in which the PI teachers the elementary methods course.

Justification: The PI clearly associates the curricular and assessment materials within the policy environment of the development and testing of effective materials, with a specific emphasis on the improvement of current elementary and middle school science curricular materials to address 'deep' ideas in science and to incorporate specific elements of scientific inquiry (explanation and data analysis).

Research Plan: The PI refers to results of prior research and development structures for both the early grade curricular material and the assessment development. The iterative design and testing of both sets of materials is characterized with a longitudinal design. The plan identifies multiple forms of evidence on the implementation of the curricular materials and initial descriptions of how the data will be used to inform improvement cycles. Reference is made to prior engineering design principles in curricular development and the connection of assessment development to the *Principled Assessment Design for Inquiry* (PADI) design principles, a foundational process of an assessment system design and validation. The quasi-experimental study with a treatment and comparison group is described. The integration of student learning with teacher science knowledge and skills in pedagogy are included with a focus on the development of teacher preparation and in-service training materials that are developed in parallel with the student materials.

Findings: The students' content knowledge is an insufficient measure for science learning. Integrating content with evidence-based explanations (what the PI calls "middle" knowledge) contributes to a more sophisticated and nuanced understanding of students' science thinking and the role that scaffolding can play in improving students' outcomes.

Publications:

Gotwals, A.W., Songer, N.B., and Bullard, L. (2012) A. Alonzo and A.W. Gotwals (Eds). Assessing Students' Progressing Abilities to Construct Scientific Explanations. In *Learning Progressions in Science,* Rotterdam: Sense Publishers.

Songer, N.B. and Gotwals, A.W. (2012) Guiding explanation construction by children at the entry points of learning progressions. *Journal for Research in Science Teaching, 49,* 141-165.

Research Type: Efficacy Study

Summer School and Summer Learning: An Examination of Selection, Implementation, and Program Effects in a Multiyear Randomized Trial

Keith Zvoch, PI

http://ies.ed.gov/funding/grantsearch/details.asp?ID=813

Purpose: The purpose of this study funded by IES is to conduct a rigorous research study of a multiyear summer school literacy program initiative delivered to kindergarten, first, and second grade students identified as at-risk for future reading difficulty and aimed at closing the performance gap between strong and struggling readers and to ensure that struggling readers gain the skills requisite to meet reading proficiency targets.

Justification: A summer school intervention for kindergarten and first-grade students at moderate risk for reading difficulties will lead to significant improvements in reading achievement immediately following the summer school program. A school-district-sponsored summer school literacy program was implemented four days per week for 3.5 hours a day over a five-week period at a single site. The curriculum was aligned with the "big ideas" and best practices from the National Reading Panel (2000) and focused on phonemic awareness and alphabetic understanding. Each kindergarten and first-grade class included approximately 20 students and four teachers. The classes began with daily lessons in a large group setting, which included checking homework and delivering seat work packets, followed by literacy instruction provided to small groups of three to five students with similar skill levels.

Research Plan: Students completing kindergarten or first grade were identified as having moderate risk for reading difficulties based on the Nonsense Word Fluency (NWF) subtest from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (for those completing kindergarten) or the Test of Oral Reading Fluency (TORF) (for those completing first grade). After identifying these students, half were randomly offered the summer school intervention, while the remaining students were not offered an opportunity to participate in the summer program. Just over half of the kindergarteners and 65% of the first graders who were offered a spot actually attended the summer program. Fidelity of implementation was monitored through a combination of classroom observation, student attendance, homework completion, and a survey of treatment and control students regarding the summer's activities given at the start of the next school year.

Reading achievement was measured for kindergarteners by the NWF of the DIBELS. Reading achievement for first-graders was measured by the TORF. The analysis compared gains in reading achievement between spring and the following fall for students in the groups to which they were originally assigned (regardless of whether they attended the program), including 46 kindergarteners and 47 first-graders.

Findings: Immediately following the intervention, students who were randomly offered admission to the summer school program scored statistically significantly higher on the reading assessments for both kindergarten and first grade (WWC-calculated effect sizes of 0.69 and 0.61, respectively).

Publications:

Zvoch, K., & Stevens, J. J. (2013). Summer school effects in a randomized trial. *Early Childhood Research Quarterly*, *28*(1), 24-32.

Zvoch, K. (2012). How Does Fidelity of Implementation Matter? Using Multilevel Models to Detect Relationships Between Participant Outcomes and the Delivery and Receipt of Treatment. American Journal of Evaluation, 33(4), 547-565.

EFFICACY RESEARCH EXAMPLE 2

The Evaluation of Enhanced Academic Instruction in After-School Programs

Alison Rebeck Black PI

(http://ies.ed.gov/pubsearch/pubsinfo.asp?pubid=NCEE20094077)

Purpose: The primary purpose of this study funded by IES was to determine the efficacy of structured academic instruction in math to students in grades two to five during their after-school time. However, the project also included a design and development effort to identify and adapt the existing mathematics curriculum that would be used in the after school setting.

Justification: In many schools, high proportions of students fail to achieve proficiency in core subjects, including math, English and language arts, and science. This study was commissioned by IES in response to theoretical and empirical evidence that a structured, academically focused after-school intervention could be an effective way to improve academic achievement. Drawing on the evidence from exploratory and design and development research, it was decided that the study would test the effectiveness of existing curricula materials that were adapted for use in an after-school setting. The adapted curriculum was delivered four days per week during the first 45 minutes of the typical two- to three-hour after-school program. The control condition was the same afterschool program with typical homework help or tutoring during those first 45 minutes of the program day.

Research Plan: Harcourt School Publishers was competitively selected to adapt its existing instructional materials for use in after-school programs. Harcourt built the after-school program around five mathematical themes or strands: numbers and operations, measurement, geometry, algebra and functions, and data analysis and probability. The 45 minute periods were constructed to mirror a gym exercise session, with a short group activity ("the warm-up"), followed by 30 minutes focused on skill building ("the workout"), and a final small-group activity to complete the session ("the cool-down"). The model includes games to build math fluency, hands-on activities, and projects, as well as computer activities for guided instruction, practice, or enrichment. Implementation was supported by strategies related to staffing, training, and technical assistance, and attendance. Thus, the evaluation was an efficacy test of an enhanced after-school program that packaged several elements: an adapted curriculum, certified teachers, small class sizes, teacher support, and attendance incentives.

Following a pilot year in 2004-2005, the model was refined and then implemented in the evaluation sites during the 2005-2006 and 2006-2007 school years, when the enhanced program instruction was offered in 25 after-school centers. Students attending the after school centers were randomly assigned to the treatment or the business as usual control condition in each of the program years. The study included both implementation research and impact analyses.

Findings: The study found that one year of enhanced math instruction produced positive and statistically significant impacts on student achievement representing approximately one month's worth of extra math learning. Two years of the enhanced program produced no additional achievement benefit beyond the one-year impact and the program impacts did not differ significantly between the first and second year of program operations.

Publications:

Black, A.R., Soumers, M.-A., Doolittle, F., Unterman, R., and Grossman, J. B. (2009). *The evaluation of enhanced academic instruction in after-school programs: Final report* (NCEE 2009—4007). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. .

Black, A.R., Doolittle, F., Zhu, P., Unterman, R., and Grossman, J.B. (2008). *The evaluation of enhanced academic instruction in after-school programs: Findings after the first year of implementation* (NCEE 2008-4021). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of education Sciences, U.S. Department of Education.

Research Type: Effectiveness Study

Working with Teachers and Leveraging Technology to Scale Opportunities to Learn More Complex and Conceptually Difficult Middle School Mathematics

Jeremy Rochelle, PI

(http://www.nsf.gov/awardsearch/showAward.do?AwardNumber=0437861)

Purpose: The project funded by NSF sought to replicate and extend a prior study of SimCalc, an interactive computer environment for mathematics instruction in grades 7 and 8. It was conducted in three distinct sites in Texas.

Justification: There is considerable need for students in grades 7 and 8 to increase their mathematical knowledge, particularly in algebra. Additionally, technological instructional tools have been lauded for potential to help students learn mathematics, in part through providing individualized feedback that is responsive to each student's performance in the instructional environment. There is a need to understand whether such technological tools can be effective in supporting instruction in a variety of settings and with a variety of students. Prior results from an efficacy study of SimCalc had shown significant effects on teachers' mathematical knowledge for teaching and on students' mathematics achievement, which demonstrated promise for the technology when replicated at multiple sites.

Research Plan: The project addressed the following research questions: Can a wide variety of teachers use an innovative technology to create opportunities for their students to learn complex and conceptually difficult mathematics? Which teaching practices positively affect students' learning with these innovative materials, and can teachers improve student learning by implementing a focused subset of these best practices? Do student gains persist after the reduction of the presence of the research and development team?

The study consisted of four parts. (1) A random-assignment experiment with about 100 seventh-grade teachers (1621 students), in which some teachers immediately began using SimCalc and others began a year later. This latter group served as a control in the first year. (2) a random-assignment experiment with about 60 eighth-grade teachers, which replicated the seventh-grade experiment but with regional leaders providing the teachers with all professional development through a typical train-the-trainers model. (3) Data collection from the seventh-grade teachers to measure implementation as the research and professional development support were reduced in later years. (4) Case studies in seventh- and eighth-grade classrooms aimed at providing more detailed qualitative data to support interpretation of experimental results and specifically to examine how the innovation interacts with student diversity.

Findings: The study found that SimCalc enables a wide variety of teachers in a diversity of settings to extend students' learning to more advanced mathematics. For the seventh-grade study, the seventh-grade quasi-experiment (delayed treatment teachers across years 1 and 2), and the eighth-grade study, the main effects were statistically significant and showed that students in the treatment group (or year 2) learned more than students in the control group (or year 1). HLM analyses revealed main effects with student-level effect sizes of .63, .50, and .56, respectively. In both the seventh- and eighth-grade experiments, these student learning effects, particularly on the advanced portions of the tests, were robust across demographic groups (i.e., gender, ethnicity,

teacher-rated prior achievement, geographic region, and campus poverty), despite marked group differences at pretest.

Selected Publications:

Roschelle, J., Shechtman, N., Tatar, D., Hegedus, S., Hopkins, B., Empson, S., Knudsen, J., & Gallagher, L. (2010). Integration of technology, curriculum, and professional development for advancing middle school mathematics: Three large-scale studies. American Educational Research Journal, 47(4), 833-878.

Shechtman, N., Roschelle, J., Haertel, G., & Knudsen, J. (2010). Investigating links from teacher knowledge, to classroom practice, to student learning in the instructional system of the middle-school mathematics classroom. Cognition and Instruction, 28(3), 317-359.

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: Advanced designs for teaching and learning. In M. J. Jacobson & P. Reimann (Eds.), *Designs for Learning Environments of the Future: International Perspectives from the Learning Sciences* (pp. 233-262). New York: Springer.

Roschelle, J., Singleton, C., Sabelli, N., Pea, R., & Bransford, J. D. (2008). Mathematics worth knowing, resources worth growing, research worth noting: A response to the National Mathematics Advisory Panel report. *Educational Researcher,* 37(9), 610-617.

Tatar, D., Roschelle, J., Knudsen, J., Shechtman, N., Kaput, J., & Hopkins, B. (2008). Scaling up innovative technology-based math. *Journal of the Learning Sciences*, 17(2), 248-286.

Roschelle, J.,Tatar, D., Schectman, N., & Knudsen, J. (2008). The role of scaling up research in designing for and evaluating robustness. *Educational Studies in Mathematics*, 68, 149-170.

Research Type: Scale-up Study

Evaluation of the Effectiveness of the Alabama Math, Science, and Technology Initiative (AMSTI)

Dennis Newman, PI

(http://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_20124008.pdf)

Purpose: The Alabama Math, Science, and Technology initiative (AMSTI) was a response by the Alabama State Department of Education to a perceived need to improve math, science and technology education state-wide. A 38-member commission developed the design for the intervention following an extensive review of the literature and consultation with content and pedagogical experts. Based on evidence from several small, matched comparison studies which showed evidence suggesting that AMSTI may improve math achievement, the state expanded AMSTI over several years and, by 2009, about 40 percent of Alabama's 1,518 public schools were designated as AMSTI schools at an annual cost to the state of $46 million a year. The present study was intended to provide credible evidence as to whether AMSTI was effective when scaled to serve this large number of schools, not only in improving math outcomes, but also in improving science outcomes. In addition, the study was designed to learn about the mechanisms through which AMSTI may operate to affect student outcomes.

Justification: The AMSTI theory of action posits that in order to improve student achievement, teacher instructional strategies should include higher levels of hands-on, inquiry-based instruction. The three components of the program that foster this type of instruction are comprehensive professional development delivered through a 10-day summer institute and follow-up training during the school year; access to program materials, manipulatives, and technology needed to deliver hands-on, inquiry-based instruction; and in-school support by AMSTI lead teachers and site specialists who offer mentoring and coaching for instruction. The full program is delivered over the course of two years. In each region, AMSTI site specialists partner with a local university or college. ALSDE oversees the professional development and implementation of the program.

Research Plan: The research plan included a primary confirmatory analyses address the effect of AMSTI on student achievement in mathematics problem solving and science after one year. The primary research question looks at whether the intervention had an effect on mathematics problem solving or science knowledge. The secondary research question addresses the effect of AMSTI on classroom practices, which are the mediating link between the intervention components and student achievement.

The study took advantage of ALSDE's rollout of AMSTI to specific regions during the study years. To participate in the study, schools must have housed at least one grade between grades 4 and 8, and at least 80 percent of a school's mathematics and science teachers must have agreed to participate. From the eligible schools that applied to the program, researchers made a purposeful effort to select a sample that was representative of the population of schools in the regions involved. Pairs of similar schools were selected from the pool of applicants based on similarity in mathematics achievement, the percentage of minority students, and the percentage of students from low-income households. Within each pair, schools were randomly assigned either to the AMSTI condition, in which teachers received AMSTI training and program materials, or to the control condition, in which teachers used their existing mathematics and science programs.

The sample included 82 schools, with about 780 teachers and 30,000 students in grades 4–8. Data were collected at multiple levels. Sources included classroom rosters, student achievement and demographic data, professional development training logs and observations, professional development teacher surveys, interviews with teachers and principals, classroom observations, and web-based surveys of teachers and principals.

Findings: AMSTI teachers were more likely to have participated in summer professional development than were control teachers, they reported having greater access to materials than did control teachers, and they were more likely to receive in-school support than were their control counterparts. The effect of AMSTI on student achievement in mathematics after one year, as measured by end-of-the-year scores on the Stanford Achievement Test Tenth Edition (SAT 10) mathematics problem solving assessment of students in grades 4–8, was 2.06 scale score units, which is equivalent to 28 days of additional student progress over students receiving conventional mathematics instruction. However, there was not a statistically significant effect on science achievement at the end of one year.

Appendix B: Common Guidelines, by Research Type

Table B- 1: Foundational Research

Purpose	The purpose of Foundational Research is to advance the frontiers of education and learning; develop and refine theory and methodology; and provide fundamental knowledge about teaching and/or learning. Foundational Research studies may examine phenomena without establishing an explicit link to education outcomes.
Justification Guidelines for Foundational Studies	
Policy and/or Practical Significance	The proposed project should address important research problems or questions related to education and learning. Although the project should have clear potential implications for policy and/or practice, a direct relationship to student outcomes is not required.
Theoretical and Empirical Basis	The proposal should outline the theoretical and empirical bases for the project and, if relevant, identify why it is necessary to develop new theory, explore learning constructs, or create more useful conceptual frameworks. The proposal should describe whether and how the project will identify or explore important new constructs in education and learning, extend understanding of current constructs, expand understanding of relationships among the constructs under investigation, and/or extend research methodologies appropriate to advancing the evidence base to support improved policy or practice.
Evidence to Be Produced by Foundational Studies	
Project Outcomes	The project's potential outcomes should include advances in theory, methodology, and/or understandings of important constructs in education and learning. Foundational Research studies should have appropriate methodological rigor such that, upon completion, the project's findings could serve as the basis for future studies.

Research Plan	The proposal should clearly define (1) key conjectures or hypotheses, research questions, and research objectives that derive from the theoretical and empirical basis of the study; (2) a detailed description of the study design, including but not limited to a description of the population(s) of interest; (3) sampling or selection methods, and the expected sample size; and (4) methods to be used for data analysis. For studies that include hypothesis testing, the proposal should discuss the minimum relevant mean difference or relationship between variables and sample size required to ensure adequate statistical power to detect true differences or relationships of this magnitude or larger. For qualitative studies or study components, proposals should provide a rationale for the sample size and selection plan. For studies that analyze secondary data, the proposal should describe the source and availability of data and the sequence of modeling planned. For studies that collect primary data, the proposal should describe the instruments and protocols that will be developed and used, provide initial evidence from literature to support assumptions that guide the sample design, and describe strategies for ensuring validity and reliability of the outcome measures, and discuss strategies for triangulation of findings. Finally, all proposals should include explicit plans for data management and analysis, including statistical models and/or procedures for analysis of text, video, or observational data, data curating and sharing, and for dissemination of findings.
External Feedback Plan	The project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator, and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings.

Table B- 2: Early-Stage or Exploratory Research

Purpose	The purpose of Early-Stage or Exploratory Research is to investigate approaches to education problems to establish the basis for design and development of new interventions or strategies, and/or to provide evidence for whether an established intervention or strategy is ready to be tested in an efficacy study. Early-Stage and Exploratory Research should establish initial connections to outcomes of interest. Studies in this genre should support the development of a well-explicated theory of action that can inform the development, modification, or evaluation of an intervention or strategy. They should build on existing research and theory to examine issues such as: (1) Associations between (a) education or learning outcomes and (b) malleable factors (that is, factors that are alterable, such as children's behaviors; technologies; education programs, policies, and practices) and; (2) Factors and conditions that may mediate or moderate the relationship between (a) education or learning outcomes and (b) malleable factors; and (3) Opportunities for new interventions or strategies, and challenges to their adoption, with the goal of informing policy, practice, and future design or development.
Justification Guidelines for Early-Stage or Exploratory Research	
Policy and/or Practical Significance	The proposal for the project should provide a clear description of the practical education problem or issue that will be the study focus and a compelling rationale for studying the problem. The project should provide a compelling case that the proposed research will generate important knowledge to inform the development, improvement, or evaluation of education programs, policies, or practices.
Theoretical and Empirical Basis	The proposal should detail a strong theoretical and empirical rationale for the project. To the extent possible, an empirical rationale should be included with citations of supporting evidence. When a study of an existing intervention or strategy is proposed, there should be a compelling explanation of why this intervention should be studied through Early-Stage or Exploratory Research rather than through Efficacy Research.

Evidence to be Produced by Early-Stage or Exploratory Studies	
Project Outcomes	The project's outcomes should include (1) Empirical evidence regarding (a) the malleable factors' association with education or learning outcomes and/or (b) evidence on whether the factors and conditions moderate and/or mediate the relations between the malleable factors and the learner outcomes; (2) (a) A well-specified conceptual framework that supports a theoretical explanation for the link between the malleable factors and the education or learning outcomes and/or (b) a theoretical explanation for the factors' and conditions' moderation and/or mediation of the relationship between the malleable factors and learner outcomes; and (3) A determination based on the empirical evidence and conceptual framework of whether there is a basis for pursuing a Design and Development Project or an Efficacy Study or whether further Foundational, Early-Stage, or Exploratory Research is needed before proceeding to efficacy or effectiveness testing.
Research Plan	The research plan should include the following: (1) a set of hypotheses or research questions that are derived from the theoretical and empirical rationale for the study; (2) a detailed research design that is appropriate for the hypotheses or research questions; (3) a justification for the proposed research context and sample; (4) a description of the data sources if secondary analyses are proposed ; (5) a detailed description of data collection procedures and instruments, as well as evidence of and strategies for ensuring reliability and validity; (6) if applicable, a plan to study the opportunities for interventions (i.e., programs, policies, practices or technologies) to address education and learning challenges; and (7) a detailed description of data analysis procedures and the reporting plan.
External Feedback Plan	The project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings.

Table B- 3: Design and Development Research

Purpose	The purpose of Design and Development Research is to develop new or improved interventions or strategies to achieve well-specified learning goals or objectives, including making refinements on the basis of small-scale testing. Typically this research involves four components: (1) Development of a solution (for example, an instructional approach; design and learning objects, such as museum exhibits or media; or education policy) based on a well-specified theory of action appropriate to a well-defined end user; (2) Creation of measures to assess the implementation of the solution(s); (3) Collection of data on the feasibility of implementing the solution(s) in typical delivery settings by intended users; and (4) Conducting a pilot study to examine the promise of generating the intended outcomes. In some cases, funders will expect all four stages to be completed within a single project; in other cases, Design and Development Projects may entail sequential projects.
Justification Guidelines for Design and Development Research	
Policy and/or Practical Significance	The proposal for the project should provide a compelling rationale that (1) specifies the practical problem the proposed intervention intends to address; (2) justifies the importance of the problem; (3) describes how the proposed intervention or strategy differs from existing practice; and (4) explains why the proposed project has the potential to improve learning or education outcomes or increase efficiencies in the education system or institutional setting beyond what current practice provides.
Theoretical and Empirical Basis	The proposal should include a strong theoretical and empirical justification for development of the proposed intervention or strategy. If the theoretical basis rests on evidence related to individual features or components, the proposal should provide a compelling rationale for how combining these features or components into a new intervention is expected to achieve intended outcomes. The proposal should include a description of the initial concept for the planned investigation, including a well-explicated theory of action or logic model. The concept and logic model should identify key components of the intervention (i.e., the ingredients hypothesized to be critical to achieving the intended results) and should describe their relationships, theoretically, and operationally.

Evidence to be Produced by Design and Development Research	
Project Outcomes	The project's outcomes should include (1) a fully developed version of the proposed design-research (including all materials necessary for its implementation); (2) a well-specified theory of action, including evidence supporting or refuting key assumptions of the intervention's original theoretical basis ; (3) descriptions of the major design iterations and the resulting evidence to support or question key assumptions about the theory of action; (4) description and empirical evidence of the adjustments to the theory of action and intervention design that resulted from design testing; (5) measures with evidence of technical quality for assessing the implementation of the intervention in an authentic education delivery setting and data demonstrating the project's success in such implementation; and (6) pilot data on the intervention's promise for generating the intended beneficial learner outcomes.
Research Plan	The research plan should describe the (1) method for developing the intervention to the point where it can be used by the intended end-users (iterative development process); (2) method for collecting evidence on the feasibility that end users can implement the intervention in an authentic education or learning setting (evidence of feasibility of implementation); and (3) method for obtaining pilot data on the promise of the intervention for achieving the expected outcomes (pilot study).
External Feedback Plan	The project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings.

Table B- 4: Impact Research

Purpose	The purpose of Impact Research is to generate reliable estimates of the ability of a *fully developed* intervention or strategy to achieve its intended outcomes. For an impact study to be warranted, the theory of action must be well established and the components of the intervention or strategy well specified. The three types of impact studies—Efficacy, Effectiveness, and Scale-up—differ with regard to the conditions under which the intervention is implemented and the populations to which the findings generalize. In addition, as the research moves from Efficacy to Scale-up, studies should also give greater attention to identifying variation among impacts by subgroup, setting, level of implementation, and other mediators. For all impact studies, descriptive and exploratory analyses should be sufficiently elaborated to determine the extent to which the findings support the underlying theory of action.
	The purpose of Efficacy Research is to determine whether an intervention or strategy can improve outcomes under what are sometimes called "ideal" conditions. For example, these conditions may include more implementation support or more highly trained personnel than would be expected under routine practice, or in contexts that include a more homogeneous sample of students, teachers, schools, and/or districts than is typical. Efficacy studies may involve the developer in the implementation of the intervention or strategy; however, the study should include reasonable safeguards for ensuring the objectivity and integrity of the study. Sometimes Efficacy studies are used to replicate previous evaluations of an intervention, but under different conditions (e.g., with a different population or using a variant of the intervention or strategy).
	The purpose of Effectiveness Research is to estimate the impacts of an intervention or strategy when implemented under conditions of routine practice. To this end, implementation should be similar to what would occur if a study were not being conducted. Importantly, an Effectiveness study should be carried out with no more developer involvement than what would be expected under typical implementation.
	The purpose of Scale-up Research is to estimate the impacts of an intervention or strategy under conditions of routine practice *and* across a broad spectrum of populations and settings. That is, Scale-Up studies should be conducted in settings and with population groups that are sufficiently diverse to broadly generalize findings. As with Effectiveness Research, Scale-up Research should be conducted with no more developer involvement than what would be expected under typical implementation.

Justification Guidelines for Efficacy Research	
Policy and/or Practical Significance	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under ideal implementation conditions with a well-defined sample, rather than under routine practice conditions with a relevant typical sample or under typical implementation conditions with a broad sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.
Theoretical and Empirical Basis	Efficacy Research should be justified by one or more of the following: (1) empirical evidence of the promise of the intervention from a well-designed and implemented pilot study (e.g., a study conducted as part of a design and development project); (2) empirical evidence from at least one well-designed and implemented Early-Stage Research study supporting all the critical links in the intervention's theory of action; (3) evidence the intervention is widely used even though it has not been adequately evaluated to determine its efficacy; or (4) if the intent is to replicate an evaluation of an intervention with a different population, evidence of favorable impacts from a previous well-designed and implemented efficacy study and justification for studying the intervention with the new target population.
Justification Guidelines for Effectiveness Research	
Policy and/or Practical Significance	The proposal for the project should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under routine practice conditions with a relevant typical sample, rather than under ideal implementation conditions with a well-defined sample or under typical implementation conditions with a broad sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.

Theoretical and Empirical Basis	Effectiveness Research should be justified by strong empirical evidence of the efficacy of the intervention, as demonstrated by statistically significant and substantively important estimates of impact, from one study that includes multiple sites or settings, or two studies that each include one site or setting, all of which meet the Evidence Standards for Impact Evaluations (Table 4) or evidence that the intervention is widely used even though it has not been adequately evaluated for efficacy.
Justification Guidelines for Scale-up Research	
Policy and/or Practical Significance	The project proposal should provide a clear description of the intervention to be tested and a compelling rationale for examining its impact. The rationale should (1) specify the practical problem the intervention is intended to address; (2) justify the importance of the problem; (3) describe how the intervention differs from other approaches to addressing the problem; and (4) explain why and how the intervention will improve education outcomes or increase efficiencies in the education system beyond current practices or interventions. The proposal should justify the choice to examine the impact of the intervention under typical implementation conditions with a broad sample, rather than under ideal implementation conditions with a well-defined sample or under routine practice conditions with a relevant typical sample. It also should describe the implementation setting(s) and population group(s) relevant to current and prospective policy or practice.
Theoretical and Empirical Basis	Scale-up Research should be justified by compelling evidence of the effectiveness of the intervention, as demonstrated by statistically significant and substantively important impact estimates from one study that includes multiple sites or settings, or two studies that include one site or setting, all of which meet the Evidence Standards for Impact Evaluations (Table 4.). In addition, there should be no overriding evidence demonstrating a negative impact of the intervention.

Evidence to Be Produced by Impact Studies	
Project Outcomes	Efficacy, Effectiveness, and Scale-up reporting should include detailed descriptions of the study goals, design and implementation, data collection and quality, and analysis and findings, for example, as outlined in the What Works Clearinghouse author reporting guidelines (http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=235). The core findings from these studies should be reliable estimates of the intervention's average impact. In some cases, it also will be possible and desirable to estimate impacts for sample subgroups defined by such characteristics as setting, population group, or cohort. Study reports should document implementation of both the intervention and the counterfactual condition in sufficient detail for readers to judge applicability of the study findings. When possible, these factors should be related descriptively to the impact findings. Study reports should discuss implications of the findings for the theory of action and, where warranted, make suggestions for adjusting the theory of action to reflect the study findings. If a favorable impact is found, the project should identify the organizational supports, tools, and procedures that were key features of the intervention implementation. If no evidence of a favorable impact is found, the project should examine possible reasons (e.g., weaknesses in the implementation, evidence that raises questions about particular aspects of the logic model).
Research Plan	The research plan should identify and justify (1) the study design used to estimate causal impact of the intervention on the outcomes of interest; (2) the key outcomes of interest for the impact study and the minimum size impact of the intervention that would have policy or practical relevance; (3) the study setting(s) and target population(s); (4) the sample, including the power it provides for detecting an impact; (5) the data collection plan, including information about procedures and measures, including evidence on and strategies for ensuring reliability and validity, and plans for collecting data on program implementation, comparison group practices, and study context; and (6) the analysis and reporting plan . Efficacy, Effectiveness, and Scale-up studies should use study designs that will yield impact estimates with strong causal validity and that, for example, could meet What Works Clearinghouse standards without reservations (http://ies.ed.gov/ncee/wwc/). Generally and when feasible, they should use designs in which the treatment and comparison groups are randomly assigned. For Impact Studies (as opposed to Design and Development Research), quasi-experimental designs, such as matched comparison groups or regression discontinuity designs, are acceptable only when there is direct compelling evidence demonstrating the implausibility of common threats to internal validity. These might include selection bias in the case of matched comparison groups, or, in the case of regression discontinuity designs, nonlinearities of

	treatment impacts over a policy relevant span around the "cut point." Ideally, the study sample size and allocation to condition should be such that the minimum true impact detectable size with 80 percent power and a 95 percent confidence interval is no larger than the minimum relevant size impact for policy or practice. If that is not the case, the proposal should provide a rationale for conducting the study despite its not meeting this standard. Primary outcome measures should include student outcomes sensitive to the performance change the intervention is intended to bring about (e.g., researcher-developed measures that are aligned with the experiences of the treatment group), student outcomes not strictly aligned with the intervention, and student outcomes of practical interest to educators and policymakers (e.g., test scores, grades, graduation or dropout rates). These outcomes should be pre-specified, have been demonstrated as reliable and valid for the intended purposes, and based on data-collection methods that have been shown to yield reliable data. The project should measure the strength and qualities of implementation (sometimes referred to as "fidelity of implementation") to address whether the intervention's impact estimates may be linked to how it was implemented. The project should measure comparison group practices and/or conditions to support a clear characterization of the contrast between the intervention and comparison condition. Proposals for Impact Studies should identify the measures, the validity and reliability of these measures, and how data will be collected on these measures. The analysis plan should specify analytic models that reflect the sample design and maximize the likelihood of obtaining unbiased, efficient estimates of average impacts and the confidence intervals around those impacts. The analysis plan should describe additional analyses conducted to explore variability in the intervention's impacts and possible implications for the theory of change. For example, these analyses could include (1) subgroup analyses (expected in Effectiveness and in Scale-up Studies); (2) exploration of co-variation in impact estimates and fidelity of implementation or intervention contrasts; and (3) evidence of possible moderator and mediator effects.
External Feedback Plan	The project should be subject to a series of external, critical reviews of its design and activities (including theoretical framework, data collection, analyses, and reporting). These review activities may entail one or more of the following: peer review of the proposed project, ongoing monitoring and review by the grant making agency's personnel, external review panels or advisory boards proposed by the project and/or the agency, a third-party evaluator and peer review of publications and conference presentations resulting from the project. The external critical review should be sufficiently independent and rigorous to influence the project's activities and improve the quality of its findings.