Building Evidence:
What Comes After an Efficacy Study?

Technical Working Group Meeting Summary
October 14, 2016

National Center for Special Education Research
National Center for Education Research
Institute of Education Sciences
550 12th Street, SW
Washington, DC 20202

Technical Working Group Members

Invited Experts

Maureen Conroy, University of Florida
Anita Zucker Endowed Professor, Professor of Special Education and Early Childhood Studies, and Co-Director of the Anita Zucker Center for Excellence in Early Childhood Studies

Fred Doolittle, MDRC
Vice President

Greg Duncan, University of California, Irvine
Distinguished Professor of Education

Mark Dynarski, Pemberton Research
Founder and President

David Francis, University of Houston
Hugh Roy and Lillie Cranz Cullen Distinguished University Chair, Director of the Texas Institute for Measurement, Evaluation, and Statistics, and Co-Director of the Center for Advanced Computing and Data Systems

Robert Granger, William T. Grant Foundation
Former President

Larry Hedges, Northwestern University
Board of Trustees Professor of Statistics and Professor of Educational and Social Policy

Carolyn Heinrich, Vanderbilt University
Professor of Public Policy and Education and Professor of Economics

Robert Horner, University of Oregon
Alumni-Knight Endowed Professor of Special Education and Clinical Sciences and Director of the Department of Special Education and Community Supports

Christopher Lemons, Vanderbilt University
Assistant Professor of Special Education

Rebecca Maynard, University of Pennsylvania
University Trustee Professor of Education and Social Policy

John Pane, RAND Corporation
Distinguished Chair in Education Innovation and Senior Scientist

John Seeley, University of Oregon
Professor of Special Education and Clinical Sciences

Jeffrey Valentine, University of Louisville
Associate Professor of Education, Counseling and Human Development

Sharon Vaughn, University of Texas at Austin
Professor of Special Education, Manuel J. Justiz Endowed Chair in Education, and Executive Director of the Meadows Center for Preventing Educational Risk

Institute of Education Sciences Staff

Elizabeth Albro, National Center for Education Research
Associate Commissioner for Teaching and Learning

Thomas Brock, National Center for Education Research
Commissioner

Jacquelyn Buckley, National Center for Special Education Research
Program Officer

Christina Chhin, National Center for Education Research
Program Officer

Emily Doolittle, National Center for Education Research
Program Officer

Joan McLaughlin, National Center for Special Education Research
Commissioner

Allen Ruby, National Center for Education Research
Associate Commissioner for Policy and Systems

Katherine Taylor, National Center for Special Education Research
Program Officer

Wendy Wei, National Center for Education Research and National Center for Special Education Research
Program Assistant

The following meeting summary was edited by IES staff for clarity and consistency. Technical Working Group members were allowed to review and comment on it, and their corrections were incorporated. The views expressed in this document reflect both individual and collective opinions of the meeting participants and not necessarily those of the U.S. Department of Education.

Meeting Summary

On October 14, 2016, the National Center for Special Education Research (NCSER) and the National Center for Education Research (NCER) in the Institute of Education Sciences (IES) convened a group of experts to gain insights and advice on how best to support the advancement of evidence beyond an efficacy study (see Appendix A for agenda). Discussion followed five topics of inquiry:

1. What should effectiveness studies accomplish? What are the challenges of initiating and completing an effectiveness study?
2. What steps are needed to encourage more replication research?
3. How can we advance our understanding of causal mechanisms and the variability of impacts?
4. How can IES training programs support the advancement of evidence beyond efficacy?
5. What is the most important thing IES can do to support the advancement of evidence once we have established efficacy?

Technical working group (TWG) members[1] made substantive suggestions about how to deepen and broaden the high-quality research supported by IES. The group called on IES to lead a "culture shift" in the evaluation of education interventions; for example, by further supporting and encouraging transparency in reporting research methods and findings, research on implementation and replication, and partnerships between researchers and practitioners. This report summarizes the IES staff presentations and participant discussions.

Effectiveness Studies at IES: An Overview

The meeting began with an overview of how effectiveness studies have been defined by IES. As described in the current Request for Applications (RFA), Goal 4, Effectiveness supports the *independent evaluation* of a fully-developed education intervention with *prior evidence of efficacy*, when implemented by the end user under *routine conditions*. In the past, Goal 4 has also been used to support scale-up evaluations of interventions with prior evidence of efficacy, the key difference being that scale-up evaluations required sufficient diversity in the sample to ensure appropriate generalizability. In total, IES has funded a small number of Goal 4 studies and there has been limited progression from Goal 3, Efficacy/Replication to Goal 4.

TWG members questioned whether the number of projects that progress from Goal 3 to Goal 4 is an appropriate measure of IES's impact. They suggested that demarcation of the two as categorically distinct types of research may be "creating a dichotomy where there is a continuum." The primary goals of IES and education researchers should be to promote schools' use of interventions and products that have an evidence base and to fill knowledge gaps where there is not an evidence base to support informed decisions about which interventions or

---

[1] The use of "TWG members" throughout the summary can refer to one or more members and does not necessarily indicate there was consensus among all the members.

products to adopt. Thus, more appropriate indicators of IES's impact on the field might be the development of such evidence and the spread of evidence-based products and interventions into schools.

TWG members encouraged IES to consider funding a study of those Goal 3 projects that did not progress to Goal 4 to better understand the barriers and potential supports for this transition. TWG members also suggested that supporting partnerships at the state and local levels (such as the IES Partnerships and Collaborations Focused on Problems of Practice or Policy program) may be a potential support. Additionally, for interventions that have evidence of efficacy under ideal conditions (i.e., conditions that provide a more controlled setting under which the intervention may be more likely to have beneficial impacts), but are not quite ready to be evaluated under routine conditions, instituting a phased approach to funding may help ease the transition to Goal 4. These studies would allow researchers to address the issues that would need to be addressed before a successful Goal 4 Effectiveness study could be conducted.

1. What Should Effectiveness Studies Accomplish? What are the Challenges of Initiating and Completing an Effectiveness Study?

During this session TWG members provided input on the evolving definition of Goal 4 (i.e., from scale-up to effectiveness) and the challenges researchers face in designing and implementing effectiveness studies.

1.1. Challenges Related to the Supply and Demand of Effectiveness Research

While efficacy, effectiveness, and scale-up studies may be best viewed as points on a continuum, there are unique challenges associated with effectiveness and scale up studies. There are not only challenges related to supplying effectiveness evidence, but there are also challenges associated with the demand for such evidence.

From a researcher perspective, effectiveness and particularly scale-up studies are risky to take on and often have uncertain payoffs. This type of research is beyond the capacity of many universities. In addition, researchers may not be motivated to do this type of research because of the length of time it takes to publish, the unwillingness of intervention developers to release control to an independent evaluator, and the lack of rewards (e.g., evidence beyond efficacy may not be needed in order to disseminate their program or practice, studies conducted over multiple years can potentially yield only one published paper).

Though there has been a continued trend toward seeking rigorous causal evidence, the distinction between levels of evidence is generally not recognized as important. Practitioners also tend to favor evidence that is established at the local level, or at least with students who they perceive as similar to their own. TWG members commended IES for the improvements to the What Works Clearinghouse (WWC) database. The ability to search for programs or practices with evidence for certain populations could be especially useful for practitioners. However, there is room for improvement; for example, the WWC does not distinguish between evidence of efficacy vs. effectiveness or make findings for subgroups of students as readily accessible as findings of overall impacts.

Overall the TWG emphasized that more work needs to be done around building a "culture of collaboration," ensuring that school partners understand why it is important to carry out the research according to established scientific practices (e.g., random assignment) and raising awareness about the importance of establishing evidence of effectiveness. The TWG also urged IES to consider ways to support researchers in conducting more effectiveness studies and increase the demand for this level of evidence among practitioners and policymakers. One suggestion was to convene local leaders in areas where effectiveness studies are taking place to process the findings from these studies and demonstrate the usefulness of evaluation findings. Another suggestion was to consider how the NCER-funded Knowledge Utilization centers could be used to better understand the demand for and use of evidence among school leaders and practitioners, and to support the aim of increasing the demand.

1.2. Do We Need a Different (or Better) Definition of "Routine Conditions"?
There are many challenges to implementing Goal 4 Effectiveness studies under "routine conditions," including those associated with using schools as laboratories. For one, researchers can only study schools, teachers, and students that agree to participate, which often leads to samples that are not representative of the population. Further, schools often impose restrictions that affect the research, such as limiting the number and type of assessments that can be administered; determining when and how to administer tests; and restricting when, for whom, and for how long professional development can be provided to teachers. It is often uncertain whether schools will retain students in the comparison condition or provide an alternative treatment after randomization. In addition, it is often difficult to establish buy-in from practitioners in the context of a grant. For example, teachers are often hesitant when asked to test a new intervention "just because some researcher got a grant to study it." Rather, they may be more concerned about implementing practices or programs that will be strong and sustained priorities at the school and/or district level.

These challenges can result in findings that are influenced more by contextual factors rather than what researchers consider desirable ways to investigate intervention impacts. Thus, some TWG members felt that researchers may need to provide more support for an intervention than would be allowed under IES's current definition of routine practice in order to conduct a true test of the intervention's impact. For instance, routine practice could be defined in a way that better aligns with the conditions that would be expected if a school or district made a calculated (as opposed to opportunistic) decision to adopt a new practice or program and was motivated to ensure that school personnel understood and successfully implemented the adopted practice or program.

1.3. Ensuring Data Integrity: Independence or Transparency?
Under the current goal structure, independent evaluators are expected to conduct Goal 4 Effectiveness studies. Some TWG members indicated that the independence of the evaluator is not as important as the transparency of the evaluation. At stake is data integrity. Presumably, bringing in a third-party evaluator eliminates or reduces the pressure to find effects in the hypothesized direction. However, an outside evaluator is not necessarily unbiased, as they may

have an interest in meeting the developer's expectations. Moreover, there may be other strategies to ensure the integrity of results; for example, by registering highly detailed research protocols prior to data collection. As one TWG member put it, "Our field has to move to registering a protocol for causal inference studies as soon as possible." Another strategy is to continue to encourage researchers in making data publicly available. The TWG urged IES to continue to support data transparency, by requiring researchers to make their final research data accessible to others even sooner than currently required (i.e., at the end of data collection as opposed to at the time of publication).

Other TWG members indicated that there is value in an independent evaluation, but that it may be better to conceptualize this as an independence of operations or functions and create a firewall between a study's implementation and evaluation. More specifically, the independent evaluator on a Goal 4 study could collaborate with the developer on the design and analysis plan, but would have sole control over the randomization, data integrity, analyses, and reporting of findings. Additionally, this conceptualization wouldn't exclude evaluators involved in previous efficacy work. For instance, it may be beneficial if the evaluator served a similar role on the Goal 3 study as they would have a deeper understanding of the program, policy, or practice.

Taking into account the value of independence and transparency, a possible approach to ensure data integrity may be to allow the developer to be the principal investigator of a Goal 4 study, but with the requirement to have an independent evaluator as defined above, to preregister the research protocol, and to make the research protocol, dataset, and all findings public, regardless of outcome.

1.4. Reframing the Purpose of an Effectiveness Study
The TWG acknowledged that effectiveness studies can be used to answer different questions, including those related to the overall impact of a practice, program, or policy implemented under routine conditions; questions about the impacts of a practice, program, or policy implemented under scaled-up conditions; and questions of what works for whom and under what conditions. They also emphasized that each of these different questions is valuable, but their value depends on the particular field of research, what is being evaluated, and the purpose of the evaluation.

In addition, often the designs utilized to answer questions of general impacts are not appropriate to answer questions about causal mechanisms and variation in impacts. The TWG offered suggestions on how best to structure effectiveness studies to answer these different questions. For instance, large-scale multi-site studies with representative samples may be a useful way to look at impact variation and better understand how the local context and local implementation affects program impacts. On the other hand, it may be more logistically feasible and/or cost effective to fund multiple, interconnected studies that seek to address these questions on a smaller scale.

Additional recommendations included focusing on building evidence of effectiveness, specifically overall impacts, earlier in the research process. Then, at a later stage, researchers could focus more heavily on questions of implementation to determine; for example, the range of conditions under which a program or practice can be implemented with fidelity and impact and the conditions under which a program or practice is sustained.

1.5. The Undervalued Role of Failed Experiments

The TWG asserted that "no funded study that is properly carried out should be lost." Moreover, studies with mixed or null effects should be viewed not as indications of research failure but as important contributions to the evidence base and opportunities to report and interpret the findings. "A well-designed study has value, even if the intervention was not effective." Sometimes negative findings should and do lead to firm conclusions about the impacts of an intervention. However, other times the conclusion is more nuanced; for example, negative effects may pertain only to a particular context. Thus, pooling findings from multiple studies can lead to a more accurate conclusion about the overall impact of an intervention and impacts within specific contexts. Obtaining mixed or null effects can also stimulate further inquiry and lead to the development of new hypotheses. IES could encourage this type of continuous improvement, by supporting researchers and organizations that are committed to understanding the reasons for mixed or null effects and what changes may lead to improved effects.

TWG members also cautioned that researchers need to be clear in describing what was evaluated and in interpreting and communicating findings, so that mixed and null effects are not misconstrued. For example, the IES-funded study of the Tennessee Pre-K initiative examined the impact of an education policy, not an educational program. However, the findings of no effect were construed as an indictment of the Pre-K programs, not the state Pre-K policy. This type of misinterpretation of what a study implies can impede the accumulation of knowledge, which is why it is critical for researchers to clearly and carefully describe their research. It is also important to situate a single study in the broader empirical history of a policy or program, as this provides a more certain picture of its effectiveness.

Current efforts to promote full reporting and dissemination of findings from all studies, including negative and null findings, involve the development of a registry for researchers to document randomized evaluations (e.g., research protocols, modifications, results). In addition, recently, some journals have been deliberate in establishing policies that do not discriminate against studies with negative or null results, including *AERA Open* and the *Journal of Research on Educational Effectiveness*. TWG members suggested that additional efforts could include an IES requirement for grantees to submit all published and *submitted* papers to the Education Resources Information Center (ERIC), so that they can be accessed by the public.

2. How Can We Encourage More Replication Research, and How Can We Overcome Barriers to Undertaking Such Research?

The topic of replication in the education sciences is complex, in part because there are different definitions of replication studies (see below). Even co-investigators sometimes disagree about

whether their study is new or a replication. This lack of clarity is compounded by the common perception that replication studies are less valuable than novel studies—a bias that could be limiting the number of replication studies that are undertaken and the amount of valuable replication research that is published. Yet replication is essential for validating prior research and advancing science.

2.1. IES Support for Replication

IES recently conducted an analysis of Goal 3 and Goal 4 studies to better identify the types of replication studies that IES has funded. It identified three distinct types of replication studies: re-analysis, direct replication, and conceptual replication. Based on this working definition, more than half of all Goal 3 and Goal 4 studies funded through 2015 have been conceptual replications; that is, a study that extends the basic parameters of the original study design in one or more specific ways (e.g., by testing the intervention with a different population, assessing its impact on different student outcomes, using a different research design or analytic approach, and/or slightly altering an aspect of the intervention or its implementation).

2.2. Finding an Optimal Balance between Replications and New Evaluations

The TWG agreed that replication is essential in building the evidence base and emphasized the importance of finding a balance of replications to new efficacy studies, especially given the strain on resources and the push to disseminate programs and practices with evidence of efficacy. One TWG member encouraged IES to prioritize re-analysis projects, as these studies are relatively inexpensive and have significant value. Using a graph showing great variability in findings from 25 different analyses of the same data set, the TWG member emphasized why investigators need to reanalyze a study's findings to determine the degree of robustness associated with different study findings before trying to replicate the research. Some findings may be quite robust (i.e., relatively insensitive to variations in models, assumptions, selection, etc.), others may not be robust, and some may not be reproducible/replicable at all.

2.3. Increasing the Visibility and Support for Replication Research

One TWG member observed: "As a field, we have not developed a framework of criteria for replication studies." The TWG called on IES to take the lead in fostering a culture shift, to appreciate the value of replication studies in building evidence. TWG members offered several suggestions for how IES could lead this effort:

- Increase funding dedicated to replication studies
- Move replications from Goal 3 to Goal 4
- Offer preferential scoring for applications that qualify as replication
- Create a new "Goal 3.5" to encourage replications that gradually release control of implementation to the schools, systematically alter characteristics of the context, utilize different analytic techniques, etc.
- Add to the RFA a more explicit definition of replication and the different types (e.g., re-analysis, direct replication, conceptual replication) and encourage applicants to clearly specify the type of replication they are proposing
- Ensure that IES-funded replications are visible (e.g., by specifying which studies are replications in the abstracts on the IES website and in the WWC database and reviews)

- Encourage researchers to build replication studies into efficacy projects (e.g., IES program officers could help investigators find opportunities for second-cohort replications, which could be added to an efficacy project at relatively low cost)

2.4. Fostering Partnerships

The TWG concurred that, especially in the current funding climate, the education science field is biased toward novel studies over replication research. The ideal ratio of novel-to-replication research may be a moving target, but IES could advocate for cost-effective replication by fostering partnerships among researchers working on similar issues. For example, IES could make connections between funded projects working on similar issues to encourage collaboration and consistency (e.g., in measures) in order to answer important questions.

TWG members recommended encouraging such collaboration and consistency so that data could be aggregated across multiple studies. IES could also consider funding a network or a Research and Development center that would bring together a diverse set of researchers to conduct replication studies of a variety of interventions.

3. How Can We Advance Our Understanding of Causal Mechanisms and the Variability of Impacts?

Effectiveness studies provide an opportunity to learn why, how, for whom, and under what conditions a program, policy, or practice works. TWG members provided feedback on the best approaches to studying the mechanisms by which an intervention works and the variability of impacts. They began by commending current efforts by IES to support work in these areas. For example, TWG members applauded IES for encouraging researchers to better measure and analyze fidelity of the treatment and the counterfactual. Given randomization is done well, the mechanisms that are driving effects have to do with the contrast; thus, it is imperative for researchers to be able to describe what is happening in the counterfactual. Additionally, the TWG noted areas in which the field needed additional guidance around studying causal mechanisms and variability in impacts and suggested steps IES could take to support more work in these areas.

3.1. Advancing the Study of Causal Mechanisms

It is difficult for investigators to develop experimental estimates of the causal relationship between mediators and outcomes; the TWG conceded that subsequent experimental trials are needed to test hypotheses arising out of mediation analyses. Nonetheless, TWG members emphasized the need to conduct better causal analyses of mediation. One TWG member emphasized the importance of using the best possible data to estimate causal mediation and using alternative approaches for robustness checking. For example, if there are three ways of estimating causal mediation, the TWG member advised researchers to do all three, and then examine relations using different data and see if the results agree. If the results agree, then that builds confidence in the findings. This underscores the need for data (e.g., from the National Center for Education Statistics (NCES) and projects funded by NCER and NCSER) to be publicly available.

The TWG suggested that IES develop a technical paper on mediation or even a short practice guide explaining what we know and possible approaches to doing causal mediation analysis. Another TWG member suggested creating a counterpart for practitioners that would describe the importance of making data publicly available. Practitioners may be more open to this if they understand the purpose and possible uses of the data.

3.2. Seeking Similarity to Understand Variability

There are two main ways to examine variation in impacts: (1) by hypothesizing where variation will occur and designing the study appropriately (e.g., randomizing at the level of expected variability) or (2) by letting variation happen and trying to model it. Either way, researchers need to measure appropriate covariates. The TWG noted that it would be helpful to have guidance about plausible candidates for covariation with impacts and common measures of these. The field has a good understanding of the demographic variables that should be included; however, the field needs additional guidance on contextual factors related to teaching and learning as these are more likely to influence treatment outcomes.

The TWG commented on the usefulness of multi-site studies in examining variability in impacts. An evaluation of Welfare-to-Work Programs conducted by MDRC was cited as a good example of a study of variability in impacts. The researchers used the same measures across sites and measured the right site-level characteristics and as a result were able to draw useful conclusions about variation in impacts. However, this type of multi-site study is not always feasible. So, in order to encourage aggregating data and pooling findings from multiple studies to answer questions that couldn't be answered in a single study, there needs to be some consistency in measurement. TWG members cited examples of initiatives by other funding agencies to provide guidance on the types of constructs to measure and to increase the consistency of measures across studies (e.g., the NIH Toolbox of neuro-behavioral measures). The TWG agreed that it would be helpful to have similar guidance and consistency in measures related to students, teachers, settings, etc.

4. How Can IES Training Programs Support the Advancement of Evidence Beyond the Stage of Proving Efficacy?

There are multiple components of training, including building technical knowledge and skills and as well as building habits and dispositions. It is worthwhile to consider how the IES training programs can train researchers to create a culture that is more accepting of replication research.

The TWG commended IES for making randomized controlled trials the gold standard and urged a continuation of this positive culture shift. Asking "What kind of research culture do we want 10 to 20 years from now, and how can IES foster the evolution of that culture?" TWG members suggested that IES:

- Establish formal networks of early career scholars through which early career investigators can learn from established researchers about how to do causal inference studies

- Provide training in project management. A lot of newer investigators are well-trained for analytics, but have not had as much exposure to designing and implementing an intervention study
- Create cross-institutional curricula for doctoral students at different universities who are studying similar issues
- Identify a specific place in the IES funding mechanism for supporting formal research syntheses
- Offer summer internships or other time-limited opportunities for pre- and post-doctoral fellows to focus on specific research activities that support efficacy and effectiveness research and are generally not fully covered in graduate school
- Offer small grants to postdoctoral fellows to replicate research findings using existing data sets and conduct research syntheses
- Develop a consortium of educational institutions to train doctoral students in order to expose them to multiple types of efficacy, replication, and effectiveness research

5. What Is the Most Important Thing IES Can do to Support the Advancement of Evidence Once We Have Established Efficacy?

The meeting concluded with individual TWG members' high-priority recommendations, which generally fell into four categories:

Rethink the Structure of Goal 3 and Goal 4
- Move replication into Goal 4
- Modify Goal 4 to encourage designs that blend effectiveness research and implementation science
- Restructure the Goal 3 to Goal 4 transition as a continuum, rather than distinctly separate stages
- Combine efficacy and effectiveness under Goal 3
- Build into Goal 3 studies a gradual release of control from researcher implementers to endogenous implementers to ease the transition from ideal to routine conditions
- Create incentives for conducting Goal 4 studies and reduce the risks (e.g., larger grants for effectiveness studies, and/or preferential treatment in scoring applications or funding decisions)

Additional Considerations for Goal 3 and Goal 4
- Require that Goal 3 and Goal 4 proposals describe how their findings – including null results – will be useful to policymakers and practitioners
- Encourage the collection of longitudinal follow-up data to determine if intervention impacts are sustained and/or to facilitate continuous improvement
- Define interventions or practices in the context of multi-tiered systems of support (MTSS) (e.g., in terms of the core features or observable elements that are directly related to student outcomes)

Facilitate More Replication Studies
- Increase the visibility of replication studies (e.g., by specifying which studies are replications in the abstracts on the IES website and in WWC products and on the website)
- Encourage re-analysis of existing data
- Shift the research culture to think more meta-analytically about research and to place more emphasis on building a body of research on an intervention strategy or approach as opposed to a single study
- Encourage re-analysis and replication by creating and continuing to enforce requirements regarding making data publicly available and publishing study protocols
- Consider ways to foster conceptual replications that systematically alter aspects of the original study design
- Leverage replications to better understand what works for whom and why

Clarify and Focus
- In the WWC database, more clearly specify the type of the study (e.g., efficacy, replication, follow-up, or effectiveness studies) and the way it was implemented (i.e., under ideal, tightly controlled conditions vs. routine conditions)
- Require that replication proposals specify the purpose of the replication and what aspects of the original study will be altered, if any
- More aggressively manage the existing portfolio of projects to encourage grantees doing similar work to collaborate and learn together
- Prioritize the development and testing of measures that can be used across many different studies, especially for certain groups of students (e.g., students with disabilities) and constructs that are difficult to measure (e.g., self-regulation), and that can be feasibly administered multiple times
- Support NCES in making it easier to link data from national datasets to other datasets

APPENDIX A. Agenda

Building Evidence: What Comes after an Efficacy Study?
Technical Working Group Meeting
Institute of Education Sciences
October 14, 2016

9 AM	**Welcome & Introductions**

Joan McLaughlin, Commissioner, National Center for Special Education Research (NCSER)
Thomas Brock, Commissioner, National Center for Education Research (NCER)

9:15   **Effectiveness Studies at IES: An Overview**
In this session, IES staff will briefly review how effectiveness studies have been defined by IES and the number of projects that have been funded. They will focus in particular on explaining the "pipeline" from Goal 3 (efficacy) to Goal 4 (effectiveness).

Presenters:
* **Elizabeth Albro**, NCER
* **Jacquelyn Buckley**, NCSER

9:45   **What Should Effectiveness Studies Accomplish? What Are the Challenges of Initiating and Completing an Effectiveness Study?**
As described in the current RFA, the effectiveness goal (Goal 4) supports the *independent evaluation* of a fully-developed education intervention with *prior evidence of efficacy*, when implemented by the end user under *routine conditions*. In the past, Goal 4 has also been used to support the evaluation of a *scaled-up* intervention with prior evidence of efficacy. In this session, we welcome the group's input on this evolving definition and the challenges researchers face in designing and implementing effectiveness studies.

Guiding Questions:
* Who values evidence of effectiveness, and for what purpose?
* Do we need a different (or better) definition of "routine practice?"
* What evidence is needed to ensure an intervention can be implemented and sustained under routine conditions?
* How important is the independent evaluation requirement? Are there other ways this might be conceptualized?
* Should IES be encouraging scale-up of interventions with prior evidence of efficacy, or is that best left to other funders?

Discussants:
- **Sharon Vaughn**, University of Texas at Austin
- **Fred Doolittle**, MDRC
- **Mark Dynarski**, Pemberton Research

Wrap-Up Commentary:
- **Carolyn Heinrich**, Vanderbilt University
- **Robert Horner**, University of Oregon
- **John Seeley**, University of Oregon

12 PM   **Lunch**

1:00   **IES Support for Replication**
There is widespread concern that there is too little emphasis on replicating studies of interventions that show evidence of efficacy. In this session, IES Staff will present information on the number and type of replication studies funded by NCER and NCSER through Goals 3 and 4.

Presenters:
- **Christina Chhin**, NCER
- **Katherine Taylor**, NCSER
- **Wendy Wei**, NCER & NCSER

1:15   **What Steps Are Needed To Encourage More Replication Research?**
This session will address the best ways to conceptualize replication research and its purposes, the barriers to engaging in replication work, and the ways that IES can support researchers in undertaking and achieving success in replication.

Guiding Questions:
- What guidance would you give IES as we try to encourage replication research?
- Are there types of replication studies that we should prioritize over others?
- What supports and incentives are needed to overcome the barriers of undertaking replication research?

Discussants:
- **Christopher Lemons**, Vanderbilt University
- **John Pane**, RAND Corporation

Wrap-Up Commentary:
- **Jeffrey Valentine**, University of Louisville
- **Maureen Conroy**, University of Florida

2:15   **Break**

2:30   **How Do We Advance Our Understanding of Causal Mechanisms and Variation in Impacts?**
Effectiveness studies provide an opportunity to learn why, how, for whom, and under what conditions a program, policy, or practice works. This session will focus on the best approaches to studying the mechanisms by which an intervention works and the variability of impacts.

Guiding Questions:
- How much emphasis should IES place on understanding causal mechanisms and variation in impacts?
- What steps can IES take to support more work in these areas?

Discussants:
- **Robert Granger**, William T. Grant Foundation
- **Greg Duncan**, University of California, Irvine

Wrap-Up Commentary:
- **David Francis**, University of Houston

3:30   **What Role Should IES Training Programs Play?**
IES supports a wide range of training programs and workshops to prepare researchers to begin careers in the education sciences, and also to build the skills of researchers already working in the field. In this session, we welcome ideas for how these programs might be modified or new programs created to address some of the challenges raised during the TWG meeting.

Guiding Question:
- How can IES training programs support the advancement of evidence beyond efficacy?

Discussant:
- **Larry Hedges**, Northwestern University

Wrap-Up Commentary:
- **Rebecca Maynard**, University of Pennsylvania

4:10   **Closing Thoughts and Recommendations**
In this session, we will do a "lightening round" to get each TWG member's one or two main recommendations for how IES can support the advancement and usefulness of effectiveness evidence. These could include issues directly under IES control and issues that IES may not have control over but may be able to influence.

Guiding Question:
- What is the most important thing IES can do to support the advancement of evidence once we have established efficacy?

4:30   **Adjourn**