

SYNERGY ENTERPRISES, INC.

AN 8(A)-CERTIFIED, WOMAN-OWNED SMALL BUSINESS



Contract Number: ED-07-CO-0025
Final Report

SATISFACTION
EXCELLENCE
INNOVATION

An Evaluation of the Effectiveness of the Institute of Education Sciences in Carrying out its Priorities and Mission

Date:
September 30, 2008

Submitted to:
Norma Garza, Executive Director
National Board for Education Sciences
U.S. Department of Education

Natasha Boyce, Contract Specialist
Contracts & Acquisitions Mgt., Group D
U.S. Department of Education

Submitted by:
SYNERGY ENTERPRISES, INC.
8757 Georgia Avenue
Suite 1440
Silver Spring, MD 20910
Phone: (240) 485-1700
Fax: (240) 485-1709
Email: info@sei2003.com
www.sei2003.com

**An Evaluation of the Effectiveness of the Institute of Education
Sciences in Carrying out its Priorities and Mission**

**Final Report by Synergy Enterprises Inc. (SEI) and the
Center for Evaluation and Education Policy (CEEP)**

September 2008

Stephen E. Baldwin, Project Director, SEI
Patricia A. Muller, Deputy Project Director, CEEP

Theresa M. Akey, CEEP
John McManus, SEI
MacDonald Phillips, SEI
Jonathan Plucker, CEEP
Sean Sharp, SEI

With the assistance of:
Kimberley Ranney, CEEP
Stephanie Schmalensee, CEEP
Joyce Stern, SEI

This report was prepared for the National Board for Education Sciences under Contract Number ED-07-CO-0025. The project officer was Norma Garza, executive director, National Board for Education Sciences.

Suggested citation: Baldwin, S.E., Muller, P.A., Akey, T.M., McManus, J., Phillips, M., Plucker, J., and Sharp, S. (2008). *An Evaluation of the Effectiveness of the Institute of Education Sciences In Carrying out its Priorities and Mission: Final Report*. Institute of Education Sciences. U.S. Department of Education. Washington, DC.

EXECUTIVE SUMMARY

The Institute of Education Sciences (IES) was established within the U.S. Department of Education by the Education Sciences Reform Act of 2002 (ESRA), which was signed into law November 5, 2002. In 2007, Synergy Enterprises Incorporated (SEI) and its subcontractor, the Center for Evaluation and Education Policy (CEEP) were contracted by the National Board for Education Sciences (NBES) to conduct an evaluation of the effectiveness of IES in carrying out its priorities and mission using primarily pre-existing data sources. Based on the overall mission and priorities of IES, the study focused on three central questions:

Rigor: To what extent, and in which ways, has IES been successful in advancing the rigor of education research?

Relevance: To what extent, and in which ways, has IES increased the relevance and usefulness of education research?

Utilization: To what extent, and in which ways, has IES increased evidence-based decisionmaking (i.e., how is the rigorous and relevant research produced through the Institute's efforts being used in education decisions)?

To some degree, the requirement to use extant data defined and limited the extent to which these questions could be addressed. In some cases original data might be collected that might better address these questions; however, given limited time and resources, many of these methods could not be employed by the present evaluation. Due to these constraints, and based on conversations with representatives from NBES regarding priorities of the Board, as well as conversations with key stakeholders at IES regarding potential data sources, the following decisions were made with regards to focus and approach: (1) The primary focus was placed on research and evaluation endeavors of IES as opposed to dissemination activities (i.e., although dissemination and utilization are addressed, the focus in terms of time and resources was on the research and evaluation functions of IES), (2) Greater emphasis was placed on competitive grants and evaluation contracts as opposed to the activities of regional laboratories or the functions of the National Center for Education Statistics (NCES), and (3) The focus was placed primarily on examining rigor as defined by a hierarchy of study designs that recognizes experimental design as the most rigorous methodology for addressing causal questions.

SEI/CEEP conducted interviews with key IES stakeholders to determine the best possible data available to answer the primary questions. Based on the available data, more specific evaluation sub-questions

Executive Summary

related to rigor, relevance and utilization were developed and mapped to the existing data sources. The only original data that were gathered within the scope of this evaluation were a limited number of key stakeholder interviews to supplement extant data. CEEP/SEI conducted interviews with key stakeholders from the following organizations and associations: American Educational Research Association, American Psychological Association, National Academy of Sciences, Council of the Great City Schools, Knowledge Alliance, and the National Sorority of Phi Delta Kappa. The revised evaluation questions and evaluation plan received approval from both the Evaluation Subcommittee of the NBES Board and the full NBES Board.

RIGOR

There are several general findings related to the primary evaluation question: To what extent, and in which ways, has IES been successful in advancing the rigor of education research? First, the emphasis on and attention to rigorous methodology, particularly randomized controlled trials (RCTs), is clearly more prominent within IES than it was within its predecessor, the Office of Educational Research and Improvement (OERI). Clear examples of the focus IES has placed on rigorous methodology are evident from the structure used for its grant programs that includes two goals focused specifically on using rigorous methodology to measure efficacy and effectiveness, and the priority placed on the What Works Clearinghouse (WWC). In addition, the fact that demand has exceeded capacity for the summer institutes on cluster randomized trials for both 2007 and 2008 suggest that education researchers understand the importance of RCTs in IES's research agenda.

Second, there has been a sharp increase in the number of RCTs being conducted within IES as compared to OERI. For example, whereas 32 percent of funded projects addressing causal questions used RCTs just prior to the establishment of IES in 2001, 82 percent to 100 percent of NCER new research and evaluation projects addressing causal questions used RCTs in the years following the establishment of IES. In addition, 24 large IES-supported evaluation studies using rigorous methodology are currently underway as opposed to just one such evaluation study in 2000 under the support of OERI.

Third, analysis of National Center for Education Research (NCER) and National Center for Special Education Research (NCSER) efficacy and effectiveness funded proposals for Fiscal Year (FY) 2004 through 07 on 10 dimensions of high quality research designs suggests that these IES studies have a high potential for generating rigorous and valid evidence of effectiveness. Although data are not yet available for the vast majority of these studies, the SEI/CEEP analysis indicates that increasing percentages of

funded efficacy and effectiveness proposals have included these dimensions of high quality research. (However, the fidelity with which these designs are being implemented cannot yet be determined.)

Finally, IES has placed a strong emphasis on increasing the capacity of the field to conduct rigorous research. To date NCER has funded 242 predoctoral fellows (2004 through 2008) and 30 postdoctoral fellows (2005 through 2008); and in July 2008 NCSER awarded five new grants for postdoctoral special education training fellowships. In addition, IES has recently begun implementing training institutes and seminars to increase researchers' skills and capacity in conducting rigorous education research (i.e., cluster randomized trials, evaluating state and district level interventions and single-case design). Demand has exceeded capacity for the two-week intensive summer institute trainings on cluster randomized designs for both 2007 and 2008. In 2007 there were almost 6 times as many applicants as participant openings and in 2008 demand was slightly more than twice the capacity for the training, suggesting that there is substantial interest from the field in increasing capacity related to rigorous methodology. What remains unknown regarding these IES initiatives is the extent to which they are effective in increasing the output of rigorous education research. For example, although preliminary data indicates that 80 percent of the persons who have completed their predoctoral fellowships are employed in research positions of some type, what remains unknown is the extent to which these interdisciplinary fellows actually pursue a research agenda related to education, and the extent to which these fellows will contribute rigorous research to the field of education.

Additional findings related to the two primary areas that the evaluation addressed with regards to rigor include the following:

Quantity and Quality of Rigorous Education Research

To what extent do the research and evaluation studies currently funded by IES meet the highest quality standards related to rigor?

- FY 07 data for a new Government Performance Results Act (GPRA) measure indicate that 100 percent of new studies of efficacy and effectiveness funded by NCER employ research designs that meet evidence standards of the WWC (target was 90%).
- IES's Program Assessment Rating Tool (PART) program performance data indicate that IES is currently meeting its targeted goals in interventions demonstrating positive effects in reading and writing and enhancing teacher characteristics (6 in 2007, and 3 in 2006 respectively), and exceeding targeted numbers of interventions in mathematics and science interventions (target of 3 in 2007,

Executive Summary

actual of 4). In addition, the number of interventions increased between 2006 and 2007 for each of the three content areas.

- The six interviewed stakeholders representing major education-related organizations strongly believe that IES has increased the quality of research being conducted within the field of education, and that the emphasis on rigor is significantly more pronounced within IES than it was during the era of OERI. Several stakeholders still noted the negative impacts of the strong focus on RCTs, but also stated that the position of IES related to rigorous research and RCTs has been modified over time, and is now more inclusive of other methodologies.
- More peer-reviewed publications were published from research grants funded during the first two years of IES than from research grants funded during the last two years of OERI (93 versus 45). For the 2 years of OERI data (2001 and 2002) versus the 4 years of IES data (2002 through 2005) this translates to an average of 11.3 peer-reviewed publications per year for OERI grants, and an average of 44.5 peer-reviewed publications per year for IES grants.

Capacity of the Field to Conduct Rigorous Education Research

To what extent has IES increased the number and quality of pre- and postdoctoral scientists? To what extent are pre- and postdoctoral scientists funded through IES programs likely to contribute to the quantity and quality of rigorous evidence related to education practice?

In terms of the PART benchmark data related to the targeted numbers of pre- and postdoctoral scientists IES hopes to impact through its fellowship training programs, although annual output targets were not met for 2007 and 2008, the actual numbers of individuals participating in the IES-funded research training programs were close to the targeted numbers (92% and 94% of the respective targets of 175 and 230).

Students participating in the IES predoctoral training programs appear to be of high quality. The average verbal GRE score was 618 (85th to 89th percentile); and the average quantitative Graduate Record Examination (GRE) score was 695 (68th to 72nd percentile). The predoctoral fellows have substantively higher GRE scores for both the verbal (i.e., 40 percentile points higher) and quantitative sections (37 percentile points higher) than intended education majors, as well as social science applicants and overall graduate school applicants. In terms of the quality of postdoctoral students participating in the fellowships, no extant data related to quality were available for the purposes of this evaluation.

In terms of the likelihood of contributing to quantity and quality of rigorous evidence, the following statistics provide some indication of the both the background and experiences of the predoctoral and postdoctoral fellows, as well as their potential to be academically productive: (a) During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES presenting a total of 662 refereed

conference presentations, and postdoctoral fellows self-reported 132 refereed conference presentations, and (b) During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES having a total of 126 published/in press papers (excluding conference proceedings), and postdoctoral fellows self-reported 52 published/in press papers.

The preliminary data related to postfellowship employment suggest that the PART target of 40 graduates engaged in research by 2009 is likely to be successfully met. In mid-2008 it appears that approximately 38 fellows of the targeted 40 are already engaged in research. However, the preliminary data do not indicate whether or not the research is specific to the field of education.

- During the past 2 years, IES has also instituted and/or funded the following trainings and information sessions aimed at increasing the capacity of the field to conduct rigorous education evaluation: 2-week Summer Research Training Institute on cluster randomized trials attended by a total of 60 participants (2007 and 2008), 1-day workshop on Evaluating State and District Level Interventions (2008) attended by 121 participants, and a 2-day IES Research Training Institute on single-case design was sponsored by NCSER (2008) attended by 39 participants.

To what extent do NCES trainings increase the capacity of education researchers to conduct rigorous education research and evaluation?

NCES conducted a total of 52 trainings on its various databases between 1999 and 2007; and the data indicate that trainings were offered for substantially more NCES databases after the creation of IES than during OERI. For example, in 1999 and 2000 there were trainings offered for only two and four databases respectively, whereas after the creation of IES the numbers of trainings consistently ranged between seven and nine per year.

On average 98 percent of trainees rated the overall quality of training as “good” or “excellent” across all surveyed years. Moreover, in 7 years, a trainee rated seminar overall quality as “poor” in only two instances (of a possible 1,318).

- In terms of potential impact on the capacity of the field to use NCES databases to conduct rigorous research, across all years a minimum of 90 percent of NCES training participants stated that they planned to use NCES datasets in the future. Approximately one-half of these participants between 2004 and 2007 had previously used a least one NCES database. However the vast majority of these same participants (between 77% and 86%) had not previously published journal articles, doctoral research, books or reports using NCES databases. Unfortunately, no data are available regarding actual usage versus intended usage or plans for using NCES databases.

RELEVANCE

Based on the established goals and priorities of IES, the evaluation also focused on the impact of the Institute on the relevance, usefulness and timeliness of education research. More specifically, the evaluation addressed the following primary question: To what extent, and in which ways, has IES increased the relevance and timeliness of education research? The six interviewed stakeholders from key education-related organizations believed that IES should get “good marks” in relevance, but also stated that they believed relevance has only more recently become a focus of IES. In terms of relevance, there are few reliable or valid data that provide insight into possible changes over time. The most current GPRA data suggests that substantial work still needs to be done in increasing the relevance of NCER and NCSER funded research: independent, external review panels found that 50 percent of funded NCSER research and 33 percent of funded NCER research is highly relevant. NCES has also historically collected data related to relevance through its customer survey. Findings generally indicate high levels of satisfaction with the relevance of NCES products, publications and services from 1997 through 2004, with levels of satisfaction similar both before and after the implementation of IES. NCES also examined differences in relevance amongst stakeholder groups, finding that although still generally very satisfied, reporters were the least satisfied with the relevance of NCES publications (i.e., 76% satisfied or very satisfied, as compared to 91% to 98% for other stakeholder groups) and policymakers were least satisfied with the ease of obtaining information from NCES (i.e., 73% indicating they were satisfied or very satisfied as compared to an average of 88.5% for all other stakeholder groups).

Relevance within the Institute was also examined in terms of the extent to which NCER and NCSER funding was aligned with the goals and priorities established by IES. In general, the Institute appears to have effectively used its overall framework for its research grant programs and its self-assessment process to identify gaps in the existing research opportunities, and has shown evidence of creating and modifying programs as needed. However, given that the most relevant and practical evidence from the perspective of practitioners and policymakers is likely to come from efficacy and effectiveness research, the absence of scale-up research within the vast majority of content areas (e.g., although the two Teacher Quality grant programs, Mathematics and Science Education and Reading and Writing, have funded a combined total of 37 grants, not a single scale-up grant has been awarded in either program; and between FY 02 and FY 07 a total of six scale-up grants have been awarded across all NCER content areas) raises some concerns in terms of relevance of the research and findings to the field. In addition, the relatively low numbers of efficacy studies in some key, long-standing content areas with relatively large research bases such as such as Reading and Writing (7 efficacy studies between 2002 and 2007) are somewhat surprising. Regardless

of whether or not this is an issue of a lack of capacity amongst education researchers to conduct this type of research, as suggested by IES, there are clear implications for the relevance of the research to the field.

Timeliness is also a factor in considering the relevance of findings and data. It is clear that NCES has embedded within its infrastructure numerous measures of timeliness, and has successfully focused its efforts on reducing turnaround time for both database releases and publications. However, a specific focus and emphasis on timeliness was not evident in the data available from the other Centers. Data related to NCER's Preschool Curriculum Evaluation Research (PCER) Initiative raises concerns about the timeliness of findings related to rigorous research. The time from final data collection for these FY 02 and FY 03 programs to the release of the final report (and individual project findings) was 3 years, with the published final report released July 2008. Given that most other programs began too recently to have final data and reports, as well as the fact that most other NCER and NCSER content areas do not include a comprehensive external evaluation component like PCER, this timeliness issue may be an anomaly. The next few years will make it more apparent whether or not the lack of timeliness was specific to the PCER program, or indicative of a broader issue with NCER funded research.

Additional findings related to the key evaluation questions concerning relevance are noted below.

Is IES providing relevant, useful, and accessible data, research and publications to various stakeholder groups? To what extent does this relevance differ among stakeholder groups?

Given concerns about the validity and reliability of the GPRA data related to the relevance of NCER funded research from 2001 through 2006, it is not possible to draw any conclusions related to relevance of NCER projects or change over time.

- The six interviewed stakeholders from major education-related organizations generally believed that IES should get “good marks” for relevance, although most persons also noted that they believed relevance has only more recently become a focus of IES. Some stakeholders also specifically noted a perceived difference in the relevance of research being funded by OERI versus IES, stating that IES appears to be better than OERI in its ability to tie research to the field.

To what extent is IES providing relevant data, and/or funding research and evaluation, that produce relevant findings as defined by the Institute's established priorities?

The mix of funded NCER research grants for the most recent years does still resemble a triangle as identified by IES to be desirable (i.e., more identification and development activities, fewer small-scale

Executive Summary

field tests, and practices at scale at the apex). However, the base of identification and development grants is wider than the overall mix of grant applications noted in the 2006 NBES Annual Report as having the desirable triangular shape. The average percentage of identification and development grants funded in 2006 through 2007 is 70.9 percent (N=56), representing a broader base for the triangle as compared to 60 percent noted for grant applications in the 2006 NBES Annual Report; and the average percentage of efficacy grants funded in 2006 through 2007 is 2.5 percent (N=2), representing a much smaller apex of the triangle as compared to 9 percent.

The majority of content areas, with the exception of High School Reform and Education Policy, Finance and Systems are making virtually no use of the identification goal that focuses on identifying programs, practices, and policies that are differentially associated with student outcomes and the factors that mediate or moderate the effects of these programs, practices and policies.

The overall pattern of NCSER funding for 2006 and 2007 is similar to NCER in terms of the majority of funded projects falling within the development category (i.e., 68% NCSER versus 64.6% NCER funded between 2004 and 2007), approximately 7 percent of funding being allocated for identification projects, and approximately 25 percent of funded grants falling within the other two aggregated categories for each Center (i.e., efficacy and scale-up).

- There are significantly more proposals funded in the IES competitions that address student achievement outcomes than under OERI (nearly 25% increase). In addition there has been a steady increase in the percentage of IES NCER studies that have addressed student achievement outcomes (a 36.5% increase) from 2004 to 2007.

To what extent is IES producing findings and data in a timely manner that ensures their relevance to current and/or pressing education issues?

For both 2006 and 2007, NCES met or exceeded its target in terms of timeliness of data releases. In addition, the percentage of NCES publications released within 18 months or less from the end of applicable data collection has increased each year, with a significant change from 2005 to 2006.

- The 2004 NCES customer survey indicates that satisfaction with the timeliness of NCES databases has increased over time from 52 percent in 1997 to 78 percent in 2004. Satisfaction with the timeliness of NCES publications has ranged from 72 percent in 1997 to 77 percent in 1999 and 2004. Satisfaction with the timeliness of NCES services remained high across the survey years, averaging 89 percent.

UTILIZATION

A third explicit goal of IES is utilization, translating the results of education research into practice. Therefore, given the explicit goal of IES in increasing utilization of rigorous research, the evaluation also addressed the following primary question: To what extent, and in which ways, has IES increased evidence-based decisionmaking (i.e., how is the rigorous and relevant research produced through the Institute's efforts being used in education decisions)? In addition, in examining utilization, the evaluation also addressed the mechanisms for education decisionmaking. More specifically, the evaluation included the following question: How, and by whom, are education decisions related to policy and practice being made in the field? What are the implications for increasing the utilization of IES research, evaluation, publications, etc.?

Similar to relevance, the six stakeholders interviewed also generally agreed that utilization was not as much of a focus for the Institute as was rigor. In fact, in terms of the three primary goals of IES (rigor, relevance and utilization), utilization uniformly received the lowest marks and most criticism from interviewed stakeholders. Valid and reliable data to confirm or disconfirm these stakeholder perceptions are not available. Data related to ERIC usage and REL calls/contacts are limited in their meaningfulness given the lack of information about who is accessing these sites/resources and for what purposes. The 2004 NCES customer survey does provide some insight into the types of data being used by various stakeholder groups. However, there is a general absence of knowledge and understanding within the field of education research about how to increase utilization of rigorous research by practitioners and policymakers (Honig and Coburn, 2008). Given this lack of understanding and knowledge, it is understandable that IES has focused primarily on increasing dissemination of information. However, without a better understanding of the ways in which rigorous research can best be integrated into policy decisions and education decisionmaking, it will be difficult for the Institute to move beyond simply increasing dissemination efforts to truly increasing utilization.

Additional findings related to the key evaluation questions concerning utilization are noted below.

To what extent, and in what ways, has IES increased evidence-based decisionmaking (i.e., to what extent is the rigorous and relevant research being used in education decisions)?

The WWC exceeded its target for website hits each year from 2003 to 2007. Annual hits increased from 1,522,922 in 2003 to 11,954,412 in 2007. Data from a web-based pop-up survey on the WWC website indicate that website visitors most frequently self-report that they plan to use the information for either K–

Executive Summary

12 classroom or home instruction or curriculum development (22% each). Respondents less frequently noted planning to use the information obtained from the WWC for policy decisions: 11 percent each noted they planned to use the information for school or district policy decisions, 4 percent noted they planned to use the information for state policy decisions, and 3 percent stated they planned to use the information for federal policy decisions. Data from the web-based pop-up survey also indicate that teachers and administrators are the most frequent users of the WWC website (23% and 19% of all respondents, respectively). In addition, approximately 12 percent of respondents included researchers. Data indicate frequent usage of ERIC, including almost 56.4 million ERIC searches conducted within a 6-month period; and an average of more than 2.7 million unique visitors per month using the eric.ed.gov website to conduct ERIC searches. Total ERIC searches have also increased over time, with a substantial increase occurring between 2006 and 2007.

Data for 2006 and 2007 indicate that there were 1.7 times more calls/contacts received in 2007 than in 2006. The increased number of calls and contacts suggests an increased use of RELs. Additional details related to the types of stakeholders making these calls/contacts, and the basic purpose of the calls/contacts, would help to draw more reliable conclusions on the relationship of these calls to education decisionmaking.

Website statistics for NCES indicate frequent visits (i.e., 11.8 million visits per year). The DAS has been accessed much less frequently, although almost half a million visits per year are reported. The NCES website clearly receives more hits than the WWC site: for NCES the average number of page views per month for 2007 was 6.38 million as compared to an average of 996,201 per month for the WWC in 2007.

- The most likely users of NCES products and data are supervisors, administrators or managers (35%) and researchers/evaluators (27%). Policymakers and reporters/media represented the smallest percentage of the distribution, with 6 percent and 2 percent respectively. External requests to NCES appear to have declined over time, with 234 total in 2006 and 162 logged in 2007. Data indicate that the decrease in overall external queries to NCES is the result of decreases in requests from the media, who make the greatest number of requests.

*How, and by whom, are education decisions related to policy and practice being made in the field?
What are the implications for increasing the utilization of IES research, evaluation, publications, etc.?*

The 2004 NCES customer survey found that across all stakeholder groups, including both NCES users and nonusers, the top three most frequent non-NCES data sources consistently included the following two sources: “your state department of education” and “other offices within U.S. Department of Education.”

For the following groups the U.S. Census Bureau was also amongst the top three: NCES-user

policymakers, NCES-user researchers/evaluators, and for both user and nonuser reporters/media. Nonuser policymakers and both user and nonuser supervisors, administrators or managers noted state or regional associations as one of the top three non-NCES data sources; and user and nonuser teachers as well as nonuser researchers/evaluators noted the American Educational Research Association as one of the top three most frequently used sources of education data.

The 2004 NCES customer survey indicates that fewer than 30 percent of all NCES data-users and fewer than 30 percent of all nonusers noted that they obtained data from Regional Educational Laboratories (RELs); and fewer than 30 percent of any single stakeholder group within either users or nonusers obtained education data from RELs. Unfortunately, the WWC was not noted specifically as a possible education data source on the survey, and therefore the survey does not provide any data related to the frequency with which various stakeholders use (or do not use) the WWC. However, since this was not an explicit purpose of the NCES customer survey, the absence of the WWC from the resources is understandable.

- The WWC was noted by the six interviewed stakeholders as the primary mechanism used by IES in its attempt to increase utilization. Some stated that the WWC was the “only real mechanism,” whereas other noted that utilization was also an intended purpose of the RELs. However, both the WWC and RELs were widely viewed by the interviewed stakeholders as not being successful in increasing utilization of rigorous research. These six stakeholders generally noted the need for additional mechanisms by which to increase utilization.

RECOMMENDATIONS

In addition to using the accessible extant data to generate these findings related to the impact of IES on rigor, relevance and utilization, the evaluation also focused on developing recommendations related to evaluating IES impact, as well as broader recommendations regarding the priorities and practices of the Institute. These recommendations include the following:

Indicators/Performance Measures. IES’s research, development, and dissemination programs recently received an effective rating, the highest score, on the Office of Management and Budget (OMB) Program Assessment Rating Tool (PART). Given that the effective rating has only been given to 18 percent of more than 1,000 programs assessed by OMB, it is clear that the Institute has established generally strong indicators and performance measures for its programs and activities. However, there are still ways in which the indicators and performance measures can be modified, or new measures developed, to further strengthen the Institutes’ ability to measure the impact of the Institute on rigor,

Executive Summary

relevance and utilization. Areas for improvement regarding specific indicators are evident throughout the evaluation report, including the following:

- (a) For the relevance GPRA indicator, the external panel used to rate the relevance of NCER-funded projects should include representatives of national educational associations (similar to the panel for NCSER) that can provide broader input than the individual principals and superintendents currently on the external review panels.
- (b) Relevance indicators should include policymakers to help provide a measure of relevance to this stakeholder group, and/or the measure should specify that it pertains specifically to relevance for practitioners.
- (c) To increase the reliability and consistency of relevance and quality measures over time, external review panels need to remain relatively stable in composition over time, and clearly delineated rubrics and standards for rating need to be established. In addition, measures of inter-rater reliability and reliability of ratings over time should be included.
- (d) Indicators related to the pre- and postdoctoral training programs need to specifically address the extent to which these individuals' postfellowship employment is specifically related to research in education, rather than simply engaged in research, particularly given the interdisciplinary nature of the fellowships. Given the resources invested into these programs, it would also be useful to collect longitudinal data related to the area of employment/research and research productivity of pre-and postdoctoral fellows. Similar data gathered from participants in the intensive summer training institutes (e.g., quantity and quality of rigorous educational research conducted prior to training and postinstitute) might also provide useful comparative data related to the efficiency and effectiveness of these two mechanisms for increasing the capacity of the field to conduct rigorous research.
- (e) Although PART assessment data includes gathering data from 2012-2014 on the percentage of persons who consult the WWC prior to making a decision, it would be helpful to also gather such data now to provide a better understanding of the extent to which the usage of WWC changes over time.
- (f) Similar to the surveys that have historically been conducted by NCES, it would be useful to also periodically collect data from a representative sample of key stakeholders (e.g., practitioners, administrators, state and federal policymakers) regarding perceptions of quality and relevance, as well as behaviors related to utilization. Unlike data obtained from web-based pop-up surveys that only gather data from those persons already using IES products or services, this type of systematic survey would provide meaningful formative and summative data related to impact on rigor, relevance and utilization.

(g) Gathering systematic performance measure data from NCER and NCSER grantees would provide a more comprehensive and consistent measure of the quality, timeliness, relevance and utilization of the data and findings generated by these grants. Systematic data can be provided by each of the grantees; and final products could also be reviewed and rated for the quality and rigor of study implementation and findings.

(h) Data for calls/contacts received by RELs should be augmented by information on the types of inquirers and the purpose of their calls/contacts in order to provide a better understanding of the utilization of REL resources and services.

(i) WWC users/stakeholders should be surveyed about the relevance and utility of intervention reports, topic reports, quick review documents and practice guides in order to provide a better understanding of the utilization of these products and their role in education-decisionmaking.

(j) A specific focus on timeliness similar to that of NCES should be implemented by NCER and NCSER to ensure that findings from funded grants are disseminated in a timely manner.

(k) The current GPRA indicator based on the percentage of NCER funded research projects that are deemed to be of high quality is questionable in terms of reliability and validity, and a measure that is independent of the funding process itself would be more meaningful. Returning to a method of having an independent panel of experts reviewing funded proposals (such as in the 2002 GPRA data) removes the assessment of quality from the funding mechanism.

NCER and NCSER Research Grant Findings. In addition to making it difficult to assess the rigor of completed studies, the lack of systematic extant data related to findings from NCER and NCSER funded projects also decreases the accessibility to these research findings, and therefore detracts from the possible utilization of the research findings by researchers, practitioners and policymakers. To increase the likelihood of utilization, as well as increase the ability to assess rigor of methodology as implemented, IES should consider making project reports more readily accessible to the public, as well as perhaps creating mechanisms for the systematic collection of data (e.g., align reporting requirements for efficacy and effectiveness studies to meet the standards of evidence criteria set out by the What Works Clearinghouse and provide a venue for detailing changes to the proposed methodology).

Capacity of Field to Conduct Rigorous Research. Given the strong interest expressed in the intensive summer training institutes on cluster randomized trials (i.e., demand exceeded capacity) and other methodological trainings, consideration should be given to expanding these programs. Since these intensive trainings target persons already in the field of education conducting research, and persons with

Executive Summary

strong interest in applying rigorous methodology to education settings, there seems to be the potential for substantial impact with relatively minimal costs compared to programs such as the predoctoral training program. Although the impact of the predoctoral fellowship program will not be evident for at least several years given the length of time needed for these individuals to begin contributing to rigorous research in education, the relatively high costs per student are readily apparent. For example, analyses of available data indicate that the average expenditure per student by predoctoral program is approximately \$176,000, with a range of approximately \$92,000 to \$333,000 per predoctoral fellow. Current estimates indicate a maximum of 80 percent of these predoctoral fellows conduct research postfellowship, and because the programs are interdisciplinary it is possible many of these fellows will not directly contribute to education research.

The costs of the predoctoral fellowships do not indicate that these fellowships are not productive or imply that they should not be continued. Further data related to impact are still needed. But the cost data does suggest that further thought should be given as to whether or not there are other mechanisms that may more quickly and efficiently increase the capacity of the field to conduct education research, such as the intensive summer training institutes. For any alternative mechanisms for increasing capacity it will be important to develop and implement measures to examine the impact of these endeavors (e.g., number of participants who successfully receive IES funding for cluster randomized trials), as well as conducting cost-benefit analyses comparing the various mechanisms for increasing capacity of the field to conduct rigorous research.

Utilization. There is a clear and definite need in the field of education for a stronger research base related knowledge use (i.e., how to increase policymakers and practitioners use of rigorous research for education decisionmaking). There is little information currently available regarding the types of evidence practitioners, administrators and policymakers use, how they use it, and what conditions help or hinder its use. Without this knowledge, IES is likely to continue to focus on increasing access to rigorous research and the dissemination of rigorous evidence rather than employing strategies that truly increase utilization of rigorous research. Although access and dissemination are critical aspects of utilization, the research base on knowledge utilization that does exist suggests that the impact of these activities will remain minimal without a stronger understanding of knowledge utilization.

The complexities of increasing utilization are acknowledged in the IES PART long-term outcome measure that focuses on the percentage of decisionmakers surveyed from 2013 through 2014 who indicate

they consult the What Works Clearinghouse prior to making decision(s) on reading, writing, math, science or teacher quality interventions. The target set for 2013-2014 is 25 percent, noted by IES in the PART document to be an ambitious goal. In other words, the long-term goal for the primary IES mechanism for increasing utilization is only 25 percent. Granted, IES is probably correct that this goal of 25 percent utilization is ambitious given that the research base on knowledge utilization suggests that policymakers and practitioners do not simply access available data and use these data to make education decisions. This type of linear relationship between rigorous evidence and decisionmaking does not exist. Therefore, a clear and strong research agenda related to better understanding how to increase the utilization of rigorous research among education practitioners and policymakers is needed. Without such a knowledge base, the resources used to increase the rigor of education research will largely remain wasted as the rigorous research that produces findings regarding “what works” will only minimally be used in education practice or policy.

Future Evaluations. Appropriate resources, and latitude in terms of scope of work, need to be given to any future evaluations aimed at assessing the extent to which the Institute has been effective in carrying out its priorities and mission. The validity and meaningfulness of findings related to the impact of IES are substantially limited when only extant data can be used for the purposes of the evaluation. There are many meaningful and useful analyses that could be included as part of an evaluation of IES if additional resources and original data collection were allowed. For example, to measure the quality and relevance of NCER-funded research over time, a random sample of projects from each year during both OERI and IES could be selected, and subsequently subjected to blind reviews (i.e., no information on the year of the proposal) by an appropriate panel of experts using carefully constructed scoring rubrics. Also, the evaluation of the impact of IES on rigor, relevance, and utilization could be enhanced by including surveys and/or interviews with past and current NCER and NCSER grantees. Data gathered through such surveys and interviews would provide the types of data needed to more validly measure the rigor and relevance of grants, and provide needed data not currently available through IES. Surveying and/or interviewing NCER and NCSER panel reviewers would be another possible method that would provide needed data to address key evaluation questions. The requirement to use extant data for this evaluation necessitated a backward mapping process whereby accessible extant data sources defined (and limited) the evaluation questions that could be addressed. Future evaluations of the effectiveness of IES in carrying out its mission need to allow the key evaluation questions to drive the design and methodology of the study.

Page left intentionally blank.

ACKNOWLEDGMENTS

The SEI/CEEP evaluation team is grateful for the guidance received from our Project Officer, Norma Garza, executive director of the National Board for Education Sciences, and from past and present Board members, including Board Chair Robert C. Granger and the members of the Evaluation Committee: Joseph K. Torgesen (chair), Jon Baron, Eric Hanushek, and David Geary.

The team also thanks the leaders and staff of the Institute of Education Sciences for providing information for the project. In addition to the Director, Grover J. “Russ” Whitehurst, these individuals include: Elizabeth Albro, Sue Betka, Jack Buckley, Phoebe Cottingham, Caroline Ebanks, Amy Feldman, Lynn Okagaki, Elizabeth Payer, Anne Ricciuti, Morgan Stair, Brian Taylor, and Brenda Wolff.

The evaluation benefited from comments provided by members of the external advisory panel: James J. Heckman, Lorraine McDonnell, David Olds, Larry Orr and Gary Walker.

Our colleagues at SEI and CEEP contributed in ways large and small during the course of the project. Among them are SEI’s Chief Executive Officer and President, Prachee J. Devadas, our contract administrator, Uyen Nhi Nguyen, publications manager, DeLicia Ballard, and her team, and, for clerical support, Rachel Peku, Malinda Stevenson, and Winona White.

Page left intentionally blank.

CONTENTS

	Page
EXECUTIVE SUMMARY	iii
ACKNOWLEDGMENTS.....	xix
INTRODUCTION	1
BACKGROUND.....	1
EVALUATION FRAMEWORK.....	2
METHODOLOGY	4
RIGOR	9
QUANTITY AND QUALITY OF RIGOROUS EDUCATION RESEARCH	11
<i>Quality Standards for Rigor</i>	12
EXTERNAL REVIEWS OF QUALITY.....	19
<i>Rigor of Evidence from IES-Supported Efficacy, Effectiveness and Research Projects</i>	22
<i>The Potential to Produce Valid and Rigorous Evidence of Effectiveness</i>	28
CAPACITY OF THE FIELD TO CONDUCT RIGOROUS EDUCATION RESEARCH	32
<i>NCES Pre- and Postdoctoral Fellowship Programs</i>	33
<i>NCES Pre- and Postdoctoral Fellowship Programs: Likelihood of Contributing to Quantity and Quality of Rigorous Evidence</i>	37
<i>NCES Database Trainings</i>	41
<i>Other IES Trainings</i>	43
SUMMARY	46
<i>Quantity and Quality of Rigorous Education Research</i>	46
<i>The Potential to Produce Valid and Rigorous Evidence of Effectiveness</i>	47
<i>Capacity of the Field to Conduct Rigorous Education Research</i>	47
RELEVANCE.....	51
RELEVANCE	51
<i>Relevance of Information</i>	51
<i>Significance of Information</i>	58
<i>IES Priorities</i>	59
<i>Do Research Opportunities Fit the Priorities?</i>	60
<i>Is the Mix of Research Appropriate?</i>	61
<i>Is the Research Yielding Findings That Will Enhance Academic Achievement?</i>	67
TIMELINESS.....	68
<i>NCES</i>	69
<i>OMB Clearance Process and RELs</i>	73
<i>NCER and NCSEER Grants</i>	75
SUMMARY OF FINDINGS.....	76
UTILIZATION	81
UTILIZATION OF IES RESEARCH AND DATA.....	81
<i>NCEE</i>	82
<i>NCES</i>	87
MECHANISMS FOR EDUCATION DECISIONMAKING	92
<i>General Education Information Needs</i>	92
<i>Stakeholder Interviews</i>	94
RESEARCH ON EVIDENCE-BASED DECISIONMAKING	97
SUMMARY OF FINDINGS.....	98
DISCUSSION AND RECOMMENDATIONS.....	101
REFERENCES	111

List of Tables

Table		Page
1	Number of criteria met for each of the eight program evaluations that have published at least preliminary results	16
2	Brief summary of main outcomes for each of the eight program evaluations that have published at least preliminary results	17
3	Number of IES-supported interventions meeting What Works Clearinghouse (WWC) standards of evidence of effectiveness: 2005–12	25
4	Established targets and actual reported Program Assessment Rating Tool (PART) data and estimates: 2005–12	36
5	Predoctoral fellow Graduate Record Examination (GRE) scores and percentiles as compared to other groups	37
6	Numbers of applicants and participants for Institute of Education Sciences (IES) trainings: 2007–08	45
7	Established target and actual percentages of new National Center for Education Research (NCER) and National Center for Education Evaluation and Regional Assistance (NCEE) projects deemed to be of high relevance: 2001–06	53
8	Percentage of National Center for Education Statistics (NCES) customer survey respondents satisfied or very satisfied with various aspects of NCES publications and services: 2004	56
9	Percentage of National Center for Education Statistics (NCES) customer survey respondents satisfied or very satisfied with relevance: 2006–07	56
10	National Center for Education Research (NCER) funded research distribution across the various goals for each of the program content areas: 2002–07	65
11	Timeliness goals for release of National Center for Education Statistics (NCES) data: 2006–09	70
12	Number of months from the end of National Assessment of Educational Progress (NAEP) Reading and Mathematics Assessment data collection to initial release of the results: 2003–09	70
13	Percentage of survey respondents that were satisfied or very satisfied with the timeliness of National Center for Education Statistics (NCES) data files, publications and services: 1997–2004	72
14	Percentage of National Center for Education Statistics (NCES) customers that were satisfied or very satisfied with the timeliness of NCES publications by stakeholder group	72
15	Percentage of National Center for Education Statistics (NCES) customers that were satisfied or very satisfied with the timeliness of NCES data files, publications and services: 2006–07	73
16	Length of time for Office of Management and Budget (OMB) clearance process for Regional Educational Laboratory (REL) projects using Randomized Controlled Trials (RCT)	74

Table	Page
17	What Works Clearinghouse (WWC) page views and visits: 2007–08 82
18	WWC Website Survey: For what purpose do you plan to use the information you obtained from the What Works Clearinghouse website during this visit? 83
19	WWC Website Survey: In what capacity are you currently visiting the What Works website? 84
20	Total Education Resources Information Center (ERIC) searches and average number of unique visitors per month: 2007–08 86
21	National Center for Education Statistics (NCES) website statistics: 2007 88
22	Percentage of National Center for Education Statistics (NCES) product users and nonusers that report obtaining education data from various non-NCES sources, by stakeholder group: 2004 93

List of Figures

Figure		Page
1	Percentage of targeted and actual NCER new research and evaluation projects that employ randomized experimental designs: 2001–2007 (measure discontinued in 2007)	13
2	Percentage of IES-funded new research and evaluation projects that are deemed to be of high quality by external review panel: 2001–2004	20
3	Percentage of targeted and actual new research proposals funded by NCER that received an average score of excellent or higher: 2003–2007	21
4	Number of peer-reviewed publications referencing IES or OERI funded grants: 2000–2005	28
5	Percentage of proposed studies addressing prestudy criteria of correct power analysis, adequate sample size, and appropriate level of randomization: 2004–2007	30
6	Percentage of proposed studies with valid outcomes, systematic data collection, and longitudinal or follow-up measures: 2004–2007	31
7	Percentage of proposed studies addressing analysis, attrition, validity threats and baseline equivalence: 2004–2007	32
8	Distribution of current positions of employed IES predoctoral fellows who have completed Ph.D. programs: 2008	40
9	Number of NCES database trainings: 1999–2000 and 2002–2007	42
10	Distribution of NCES database trainees by occupation: 2004–2007	43
11	Distribution of NCER-funded research by goal category: 2002–2007	63
12	Distribution of NCSER-funded research by goal category: 2006–2007	67
13	Percentage of funded studies with student achievement outcomes: 1996–1997 and 1999–2007	68
14	Percentage of NCES publications released in 18 months or less: 2005–2007	71
15	What Works Clearinghouse annual website hits (in millions): 2003–2007	82
16	Estimated number of ERIC searches (in millions): 2005–2007	85
17	Estimated number of calls or contacts received by RELs: 2006–2007	87
18	Distribution of NCES product users by involvement in education: 2004	89
19	Number of external queries to NCES: 2006–2007	90
20	Distribution of external queries to NCES by organization type: 2006–2007	91

List of Exhibits

Exhibit		Page
A	Institute of Education Sciences: Goals, priorities and performance indicators noted in various documents	7

INTRODUCTION

Background

The Institute of Education Sciences (IES) was established within the U.S. Department of Education by the Education Sciences Reform Act of 2002 (ESRA), which was signed into law November 5, 2002. The work done under the current contract to evaluate the effectiveness of IES in carrying out its priorities and mission focuses on the effectiveness of IES in increasing evidence-based knowledge about what works in education and in disseminating that knowledge to policymakers, educators, parents, and the broader community of stakeholders with an interest in improving educational outcomes in the United States. The ESRA described the overall IES charter as the following:

“The mission of the Institute is to provide national leadership in expanding fundamental knowledge and understanding of education from early childhood through postsecondary study, in order to provide parents, educators, students, researchers, policymakers, and the general public with reliable information about (A) the condition and progress of education in the United States, including early childhood education; (B) educational practices that support learning and improve academic achievement and access to educational opportunities for all students; and (C) the effectiveness of Federal and other education programs”

A set of long-term research priorities aligned with this general mission was subsequently developed by Dr. Grover Whitehurst, IES director, and concurred with by the National Board for Education Sciences (NBES) in September 2005. These priorities include:

“...First, to develop or identify a substantial number of programs, practices, policies, and approaches that enhance academic achievement and that can be widely deployed; second, to identify what does not work and what is problematic or inefficient, and thereby encourage innovation and further research; third, to gain fundamental understanding of the processes that underlie variations in the effectiveness of education programs, practices, policies, and approaches; and fourth, to develop delivery systems for the results of education research that will be routinely used by policymakers, educators, and the general public when making education decisions...”¹

The general mission of the Institute as defined by ESRA, along with the long-term research priorities and goals, provide the underlying context for understanding the nature and purpose of the proposed

¹ <http://ies.ed.gov/director/board/priorities.asp> (retrieved August 22, 2008).

Introduction

evaluation. However, a variety of other extant documents (e.g., IES 2006 Annual Report, IES 2004 First Biennial Report to Congress, and Government Performance Results Act [GPRA] indicators) also provide critical data related to the explicit goals that have guided the work of the IES. Exhibit A on page 5 provides an overview of various goals, priorities, and performance indicators that have been noted as central to the functions of the Institute.

A valid and meaningful evaluation of the effectiveness of IES needs to go beyond the formal mission of the Institute to incorporate the actual purposes and goals that have guided its day-to-day work. Therefore, this evaluation is based on both the formal mission and Statement of Work (SOW), as well as a preliminary review of the other various goals and indicators noted in exhibit A. For example, although the more general IES mission is to expand education knowledge and provide reliable information to various stakeholders, the Institute has clearly focused its policies, practices, and resources on increasing the rigor, relevance, and utilization of education research. Therefore, this evaluation uses these three IES goals (i.e., rigor, relevance and utilization) as the framework for the evaluation.

Evaluation Framework

The primary goal of the proposed evaluation is to determine the extent to which IES has been effective in carrying out its priorities and mission. The key objectives of the evaluation are to:

Provide valid and reliable evidence related to the Institute's effectiveness, progress, and overall impact, within the available timeframe and scope of work;

Provide the groundwork for future evaluations and the collection of ongoing performance data that can be used to measure the Institute's progress over time; and

Provide policy and program recommendations based on the findings to enhance the ability of the Institute to carry out its priorities and mission.

As noted earlier, the evaluation achieves these objectives by focusing on three central IES goals: increasing rigor, increasing relevance, and increasing utilization of Institute research. Therefore, the key evaluation questions are organized around these three central themes. The three primary questions are:

Rigor: To what extent, and in which ways, has IES been successful in advancing the rigor of education research?

Relevance: To what extent, and in which ways, has IES increased the relevance and usefulness of education research?

Utilization: To what extent, and in which ways, has IES increased evidence-based decisionmaking (i.e., how is the rigorous and relevant research produced through the Institute’s efforts being used in education decisions)?

However, the extent to which each of these three primary evaluation questions can be addressed through the current evaluation project is limited by two significant factors: (1) the relatively short timeline and limited resources required decisions to be made related to the evaluation’s specific focus and general approach (e.g., breadth versus depth), and (2) the scope of work for the evaluation required that only pre-existing sources of data be used to address these primary questions, with the exception of limited key stakeholder interviews. In other words, in most instances there are more valid and meaningful data that could better addresses these evaluation questions. However, given the limited resources in terms of time and scope of work, many of these methods could not be employed for the purposes of the evaluation; and allowing only extant data to be used to address these evaluation questions (with the exception of limited stakeholder interview data) strongly limited the types of questions that could be validly and meaningfully addressed within the scope of the current evaluation. For example, although measures like website hits are imperfect proxies for utilization, such data were often the best or only data currently available and accessible. Due to these constraints, and based on conversations with representatives from NBES regarding priorities of the Board, as well as conversations with key stakeholders at IES regarding potential data sources, the following decisions were made with regards to focus and approach:

The primary focus was placed on research and evaluation endeavors of IES as opposed to dissemination activities (i.e., although dissemination and utilization are addressed, the focus in terms of time and resources was on the research and evaluation functions of IES).

Greater emphasis was placed on competitive grants and evaluation contracts as opposed to the activities of regional laboratories or the functions of NCES.

The focus was placed primarily on examining rigor as defined by IES’s hierarchy of study designs that recognizes experimental design as the most rigorous methodology for causal questions.

The decision to focus primarily on randomized controlled trial (RCTs) studies in the examination of IES’s impact on advancing the rigor of education research is a decision based on the limited scope and resources for the current evaluation. Increased rigor can be measured in multiple ways, and in fact for

Introduction

some of the IES Centers, it is not meaningful to limit the definition of rigor to RCTs. For example, focusing on RCTs in an examination of the impact of NCES on increasing the rigor of education research is not meaningful or possible. However, even within NCER and NCSEER there is a recognition of the value of other methodology such as regression discontinuity, and a belief that the proposed methodology needs to be aligned with the given research question (e.g., although RCTs provide rigorous methodology for addressing causal questions, other exploratory research questions might be better addressed using different methodology) and take into account the complexities of the real world implementation of the proposed study design. Therefore, given additional resources, a more comprehensive evaluation of the impact of IES in increasing the rigor of education research would also include a broader definition of rigor that takes into account the differing missions of the four IES centers and the strengths of the various methodologies for addressing different research questions. However, the limited scope and resources for the proposed evaluation necessitated a more narrow focus for the purposes of this study. Therefore, given the strong emphasis IES has placed on increasing RCTs within education, a decision was made to focus primarily on examining rigor as defined by IES's hierarchy of study designs that recognizes experimental design as the most rigorous methodology for causal questions. The decision for this focus was guided by input from members of the Evaluation Subcommittee of the NBES Board; and subsequently approved by the NBES Board as part of the overall evaluation plan.

Methodology

As noted previously, the primary evaluation questions needed to be addressed using pre-existing data sources. Therefore, in order to determine the data sources available to address the primary research questions, SEI/CEEP conducted interviews with the following individuals: Grover Whitehurst, director of IES; Sue Betka, deputy director for administration and policy; the commissioners and/or their representative for each of the four respective centers, and Anne Ricciuti, deputy director for science. Based on their knowledge of available data, more specific evaluation sub-questions related to rigor, relevance, and utilization were developed and mapped to the existing data sources. The revised evaluation questions and evaluation plan were provided, and received approval, from both the Evaluation Subcommittee of the NBES Board and the full NBES Board.

After completing the relevant security clearance processes, SEI/CEEP subsequently requested all relevant data from IES and its respective Centers. In some instances, extant data initially thought to be available were not able to be provided for purposes of the evaluation. For example, only 13 files for OERI projects that address causal questions² could be located because “Many of the files from fiscal years 1998, 1999 and 2000 have been closed over 5 years. Consistent with the records and information management directives, the files have been destroyed.”³ In other instances, data that were available had reliability or validity issues. However, extant data deemed valid and reliable were analyzed and included in the evaluation in all cases. Details related to the various sources of these data are included in each subsection of this report.

As noted, SEI/CEEP conducted key stakeholder interviews to supplement extant data. Interviews were completed with key stakeholders from the following organizations and associations: American Educational Research Association (AERA), American Psychological Association (APA), National Academy of Sciences, Council of the Great City Schools, Knowledge Alliance, and National Sorority of Phi Delta Kappa. Interviews were refused or declined by the National Education Association, the Council of Chief State School Officers, and the American Federation of Teachers; and various members on the board for the Society for Research on Educational Effectiveness did not respond to multiple requests for an interview. In addition, interviews were conducted with one House minority committee staff and one Senate minority committee staff. Despite repeated attempts at scheduling interviews with more than fifteen different legislative aides or committee members, the vast majority either declined to participate and/or did not return multiple phone calls and e-mails contacts requesting their participation.

Given the small numbers of legislative respondents, these data are not included in the reported results. However, findings from stakeholder interviews with representatives from key education-related organizations are included in the evaluation report. Interpretation is somewhat limited due to the small numbers of education-related organizations represented. However, the stakeholders included in the data represent some of the largest and most representative education-related organizations in the nation (e.g.,

² Research projects addressing causal questions are studies that examine how changing one variable affects another—such as impact, change over time, and differences in groups that are attributed to different levels of an intervention (not necessarily an RCT).

³ Source: E-mail from Brenda Wolff to Norma Garza on 6/24/08, copying SEI/CEEP staff.

Introduction

AERA and APA); and the interview responses represent these persons' perceptions of the views and opinions of their broader constituencies, rather than the individual opinions of six persons. Therefore, the data from these six interviews do provide some insight into perceptions of IES impact, particularly when interpreted within the context of other available data.

Exhibit A. Institute of Education Sciences: Goals, priorities and performance indicators noted in various documents

IES Mission from ESRA			
<p><i>"IN GENERAL: The mission of the Institute is to provide national leadership in expanding fundamental knowledge and understanding of education from early childhood through postsecondary study, in order to provide parents, educators, students, researchers, policymakers, and the general public with reliable information about (A) the condition and progress of education in the United States, including early childhood education; (B) educational practices that support learning and improve academic achievement and access to educational opportunities for all students; and (C) the effectiveness of federal and other education programs."</i></p>			
Statement of Work: Overarching Goals from Table 1 Research Questions and Data Sources			
<p>(1) IES will expand knowledge on: (a) the condition of education in the United States and comparisons with other countries, (b) practices that improve academic achievement and access to educational opportunities for all children, and (c) the effectiveness of federal and other education programs.</p> <p>(2) IES will provide information to parents, educators, students, researchers, policymakers, and the general public: (a) IES will disseminate information on the condition of education in the United States and provide comparative international statistics, (b) IES will develop delivery systems for the results of education research that will be routinely used by policymakers, educators, and the general public when making education decisions.</p> <p>(3) IES will transform education into an evidence-based field.</p>			
Institute Research Priorities (Various Documents)	Goals listed in IES Annual Report (July 2006)	Goals in the First Biennial Report to Congress (2004)	GPRA Indicators (From FY 07 Performance Budget)
<p>(1) By providing an independent, scientific base of evidence and promoting its use, the Institute aims to further the transformation of education into an evidence-based field, and thereby enable the nation to educate all of its students effectively.</p> <p>(2) In pursuit of its goals, the Institute will support research, conduct evaluations, and compile statistics in education that conform to rigorous scientific standards, and will disseminate and promote the use of research in ways that are objective, free of bias in their interpretation, and readily accessible.</p>	<p>(1) To develop or identify a substantial number of programs, practices, policies and approaches that enhance academic achievement and that can be widely deployed.</p> <p>(2) To identify what does not work and what is problematic or inefficient, and thereby encourage innovation and further research.</p> <p>(3) To gain fundamental understanding of the processes that underlie variations in the effectiveness of education programs, practices, policies, and approaches.</p> <p>(4) To develop delivery systems for the results of education research that will be routinely used by policymakers, educators, and the general public when making education decisions.</p>	<p>(1) Rigor of research.</p> <p>(2) Relevance of research.</p> <p>(3) Utilization of research.</p>	<p>(1) Evidence-based approaches (Utilization): The proportion of school-adopted approaches that have strong evidence of effectiveness compared to programs and interventions without such evidence.</p> <p>(2) Quality (Rigorous standards for education research): Percentage of new research proposals funded by the Departments NCER/NCSE that receive an average score of excellent or higher from an independent review panel of qualified scientists.</p> <p>(3) Relevance: Percentage of new research proposals funded by the Departments NCER/NCSE that are deemed to be of high relevance from by an independent review panel of qualified practitioners.</p>

Page left intentionally blank.

RIGOR

“The complex world of education—unlike defense, health care, or industrial production—does not rest on a strong research base. In no other field are personal experience and ideology so frequently relied on to make policy choices, and in no other field is the research base so inadequate and little used” (National Research Council, 1999).

Assessments of the state of education research such as the one above by the National Research Council led to the creation of IES, and the Institute’s considerable efforts to advance the rigor of education research. One of the most immediate changes realized with the creation of IES was the establishment of a new system for the scientific peer review of grant applications in FY 02. The system is similar to the process of grant application peer review at the National Institutes of Health and includes IES’s Standards and Review Office recruiting highly qualified reviewers primarily on the basis of the quality of their research, publications in scientific peer-reviewed journals, and the degree to which they are in-depth experts in the relevant research methods and content matter. The National Board for Education Sciences (NBES), as required by statute, reviewed IES’s peer-review processes and as reported in the 2006 NBES Annual Report, found the peer-review process to be of the “highest merit” and comparable to those of other federal agencies such as the National Science Foundation (NSF) and National Institute of Child Health and Human Development (NICHD). Many of the Board members who have received funding from NSF and/or the NICHD noted they were impressed with the system the Institute had implemented and were able “to validate that these processes would assure quality, objectivity, validity, and integrity in scientific publications” (NBES, 2006, pp. 10).

Given these previous findings related to the new system for the scientific peer review of grant applications implemented by the Institute, this evaluation did not specifically include an assessment of the impact of IES and the peer-review system on the rigor of education research. Instead the evaluation focused on other mechanisms the Institute has employed in its efforts to increase the rigor of education research. The overall research question was:

To what extent, and in which ways, has IES been successful in advancing the rigor of education research?

As noted previously in the introduction, for the purposes of this particular research question the focus is primarily on examining rigor as defined by IES’s hierarchy of study designs that recognizes experimental design as the most rigorous methodology for addressing causal questions. In an October 2002 Evidence-

Rigor

Based Education (EBE) presentation by Grover J. (Russ) Whitehurst at the Student Achievement and School Accountability Conference,⁴ Dr. Whitehurst noted the following levels of evidence ranging from highest quality to lowest quality: randomized trials, comparison groups (quasi-experimental design), pre-post comparisons, correlational, case studies, and anecdotes; and in an April 2003 presentation to AERA's annual meeting,⁵ Dr. Whitehurst stated that "randomized trials are the gold standard for determining what works." However, during this same presentation in which Dr. Whitehurst stated "Randomized trials are the only sure method for determining the effectiveness of education programs and practices" he also emphasized the following positions of IES:

"Randomized trials are not appropriate for all questions."

"Interpretations of the results of randomized trials can be enhanced with results from other methods."

"A complete portfolio of Federal funding in education will include programs of research that employ a variety of research methods."

"Questions of what works are paramount for practitioners; hence randomized trials are of high priority at the Institute."

The decision to focus primarily on RCTs in examining the impact of IES on increasing rigor is based on the self-stated emphasis and priority IES has placed on experimental design. However, this decision is based on the limited scope and resources for the current evaluation and does not imply that randomized trials are appropriate for all questions or that other types of research methods are not being funded by IES. Given additional resources, a more comprehensive evaluation of the impact of IES in increasing the rigor of education research would also include a broader definition of rigor that takes into account the differing missions of the four IES centers and the strengths of the various methodologies for addressing different research questions.

In terms of the available extant data, the two primary areas that this evaluation will address in this section with regards to rigor are the following:

⁴ http://ies.ed.gov/director/pdf/2002_10.pdf (retrieved September 21, 2008).

⁵ http://ies.ed.gov/director/pdf/2003_04_22.pdf (retrieved September 21, 2008).

1. Quantity and quality of rigorous education research
2. Capacity of the field to conduct rigorous education research

Each of these areas is discussed in more detail below. Accessible extant data related to both the quantity and quality of rigorous education research are provided, first addressing the extent to which the research and evaluation studies currently funded by IES appear to meet the highest quality standards related to rigor, and then the extent to which these research and evaluation studies are producing (or likely to produce) valid evidence. Next a discussion of the capacity of the field to conduct rigorous education research is provided, including data related to the various strategies employed by the Institute in its efforts to increase capacity. Finally, a brief summary of findings is included.

Quantity and quality of rigorous education research

Given the limitations on this evaluation (i.e., the need to use only accessible extant data), determining the impact of IES on increasing the number and/or quality of rigorous education evaluations is difficult. Indeed, using IES's own hierarchy of study designs that posits randomized control trials as the gold standard suggests that findings from the present evaluation about IES's impact on the rigor of education research are tentative. It is difficult to develop strong and rigorous findings given the limited availability of data and inability to design and implement a more rigorous methodology that might help better approximate causal impact.

Therefore, rather than focusing specifically on what impact IES has had on rigorous research, we focus on describing the indicators that suggest that rigor may be increasing. It should be noted that many of the extant data sources are necessarily limited with respect to understanding factors that may have led to changes in indicators—some likely due to structural, procedural, and mission changes of IES (such as clearly specifying the characteristics of studies that will and will not be funded), or due to other factors (e.g., differences in review panels across time, funding level differences). We are also not able to make direct links between program practices and policies that led to actual changes, although we can make some limited speculation about these links. Therefore, our discussion below focuses primarily on the following two questions:

To what extent do the research and evaluation studies currently funded by IES meet the highest quality standards related to rigor?

To what extent are these research and evaluation studies producing (or likely to produce) valid evidence?

Rigor

The best available data sources for addressing these questions included: (1) the percentage of IES-funded studies rated as high quality; (2) the number of IES-supported interventions meeting What Works Clearinghouse (WWC) standards for evidence of effectiveness; (3) the degree to which National Center for Education Evaluation and Regional Assistance (NCEE) evaluation study contracts reflect the standards of rigor put forth to guide IES; (4) analysis of funded proposals to assess the likelihood that they will produce valid and rigorous evidence of effectiveness; (5) comparisons to Office of Educational Research and Improvement (OERI) funded studies; (6) stakeholder perceptions of the rigor of research supported by IES, and (7) publications in peer-reviewed journals based on research from IES and OERI funded projects.

Quality Standards for Rigor

To what extent do the research and evaluation studies currently funded by IES meet the highest quality standards related to rigor?

There has been a substantial increase in the number of grants and evaluation studies funded by IES since its inception and compared to its predecessor OERI. This increase can be seen across Centers, including the National Center for Education Research (NCER), National Center for Special Education Research (NSCER), and NCEE.

NCER Research Grants. NCER primarily funds research conducted by individuals and teams of investigators at universities and other nonprofit research organizations. NCER has developed focused research competitions that target topics that are IES priorities. The number of research competitions increased from three in FY 02 to 11 in FY 07; the number of applications received increased from 226 in FY 02 to 459 in FY 07. OERI, the predecessor organization to IES, had 89 active grants funded in FY 01. IES had 265 active grants funded from the same funding line in FY 06, a roughly threefold increase.

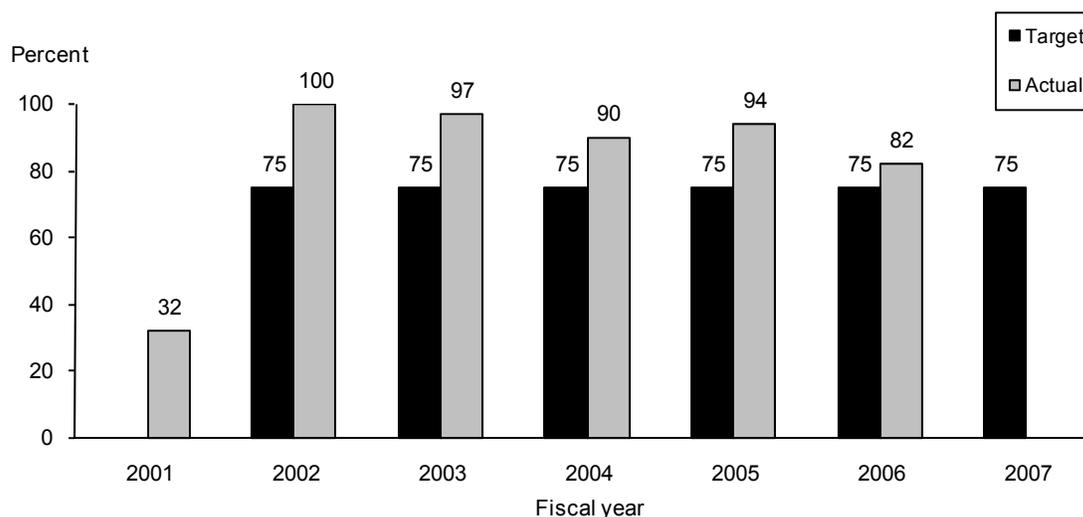
The simplest indicator of rigor is to examine the number of studies addressing causal issues (i.e., the degree to which educational interventions influence and change a variety of outcomes, particularly student learning, and classroom pedagogy) that employ RCTs. Government Performance and Results Act (GPRA) indicator data on the number of research designs that address causal research questions and use RCTs are available for FY 01 through FY 06, encompassing both OERI and IES funding years. The GPRA indicator is defined as the following:

“Of the new research and evaluation projects funded by the Department that address casual questions, the percentage of projects that employ randomized experimental designs”.

To determine the percentage of new research and evaluation projects that employ randomized experimental designs to address causal questions, this GPRA indicator included two researchers reviewing a random selection of grant proposals funded each fiscal year to determine, (1) did the principal investigator (PI) of the proposal ask a causal question?, and (2) did the PI propose a randomized experimental design to answer the causal question? A minimum inter-rater reliability of 90 percent was maintained across funding years.

Figure 1 indicates the percentage of funded studies that address causal questions that use RCTs. As noted in the figure, prior to the establishment of IES the procedure described above found that only about one third (32%) of funded projects addressing causal questions used randomized experimental designs. However, immediately after the establishment of IES, there was a drastic increase in the percentage of education research and evaluation projects addressing causal questions proposing RCTs. In the fiscal years since FY 02, percentages have remained high, ranging from 82 percent to 97 percent.

Figure 1. Percentage of targeted and actual NCER new research and evaluation projects that employ randomized experimental designs: 2001-2007 (measure discontinued in 2007)



SOURCE: U.S. Department of Education, FY 2007 Program Performance Report.

This GPRA measure was discontinued in 2007 “because it did not focus specifically on the most appropriate types of proposals, nor use the most appropriate benchmark of research design quality.” Thus, a replacement GPRA measure was developed that focused on the evidence standards of the WWC. More specifically, the GPRA measure starting in 2007 stated that:

Rigor

“Of new studies of efficacy and effectiveness funded by the Department’s National Center for Education Research (NCER), the percentage that employ research designs that meet evidence standards of the What Works Clearinghouse.”

The WWC standards include an indicator specific to the use of RCTs, but also include broader criteria for evaluating rigor (e.g., sample sizes, power analyses, attrition). For FY 07, the targeted percentage was 90 percent, and the actual percentage was 100 percent.

NCEE Evaluations. NCEE is charged with the task of evaluating the impact of programs administered by the U.S. Department of Education using “methodologically rigorous designs applied to large samples of students and schools.” To date, 24 large IES-supported evaluation studies are currently underway. These evaluations cover a range of educational programs, including early and middle-grade literacy programs, teacher preparation and professional development, math curricula afterschool academic programs, charter and magnet schools, English language learners, educational technology, and postsecondary transition programs. The number of such evaluations using rigorous methodology in 2000 under the support of OERI was one evaluation study.

Reports of impact findings are currently available online for eight of these evaluation studies. (Some studies have multiple reports.) We provide a review of the degree to which these reports of impact findings align with applicable indicators from the *What Works Clearinghouse Evidence Standards for Reviewing Studies*.⁶ Our review focused on the following ten indicators of study quality:

Did they use random controlled trials (RCTs)?

Did they use appropriate power analyses?

Were the sample sizes adequate to discern real increases in student performance?

Were the levels of overall attrition and differential attrition within standard (e.g., less than 30% overall or 5% between treatment and control groups), and were missing data addressed in impact analysis (e.g., using an intent-to-treat approach and/or estimating bias from missing data)?

Did the study demonstrate baseline equivalence between treatment and control groups, or control for any differences in the analysis?

⁶ http://ies.ed.gov/ncee/wwc/pdf/study_standards_final.pdf (retrieved August 21, 2008).

Were valid outcome measures used?

Was there consistent data collection for both treatment and control groups?

Was there appropriate follow-up (“longitudinal”) data collection to identify robustness of effects?

Did the study account for other threats to internal and external validity?

Were both positive and negative impact findings reported?

Table 1 shows the number of criteria met for each of the eight program evaluations that have published at least preliminary results. In addition, to provide a broader context, table 2 also provides a brief summary of the main outcomes for each of the eight NCEE funded program evaluations.

Random assignment to intervention or control groups is the best way to eliminate selection bias. Six out of the eight studies⁷ randomly assigned students, teachers, or schools to either an intervention group or a control group. Other methods of creating intervention and control groups included regression-discontinuity designs and equated quasi-experimental designs. In all cases but one, the reports examined issues of baseline equivalence between the treatment and control groups and found the two groups to be highly similar in each study.

An adequate sample size is necessary to detect meaningful effects of an intervention. The discussions of sample size in the design studies indicated a strong awareness of sample size and the ability to obtain meaningful results. Judging from these discussions, all eight evaluation studies appeared to have an adequate sample size, based on a power analysis that determines the minimum detectable effect size for their sample. In addition, six of the eight studies provided detailed descriptions of attrition and methods for handling it—including estimating and accounting for bias in the statistical modeling and using intent-to-treat approaches to estimate the impact for all participants in the study.

⁷ One study used a combination of RCT and quasi-experimental methods for elementary and middle school cohorts.

Table 1. Number of criteria met for each of the eight program evaluations that have published at least preliminary results

Program evaluations	RCT ¹ used	Power	Adequate sample size	Treatment of attrition and missing data	Baseline equivalence	Valid outcomes	Consistent data collection	Longitudinal or follow-up measurement	Other threats to validity	Evidence of effect
Evaluation of the DC Opportunity Scholarship program impacts after 2 years	✓	✓	✓	✓	✓	✓	✓	✓	✓	Mixed
Evaluation of Enhanced Academic Instruction in Afterschool programs: First year impact findings	✓	✓	✓	✓	✓	✓	✓	✓	✓	Mixed
Reading First Impact Study Interim Report	(²)	✓	✓	✓	✓	✓	✓	✓	✓	Mixed
The Enhanced Readings Opportunity: Early Impact and Implementation Findings	✓	✓	✓	✓	✓	✓	✓	✓	✓	Positive
Evaluation of Title 1: Final Report	✓	✓	✓	✓	✓	✓	—	✓	✓	Mixed
National Evaluation of Early Reading First: Final Report to Congress	QE ³	✓	✓	✓	✓	✓	✓	✓	—	Mixed
Effectiveness of Reading and Mathematics Software Products	✓	✓	✓	✓	✓	✓	✓	—	✓	Mixed
Third National Even Start Evaluation	✓	✓	✓	—	✓	✓	✓	✓	—	No Impact
National Evaluation of the 21 st Century Community Learning Centers Program	RCT ¹ and QE ³	✓	✓	—	—	✓	✓	✓	✓	Mixed

¹ Randomized Controlled Trial (RCT)

² The Reading First Impact study used a quasi-experimental regression discontinuity design.

³ QE = Quasi-experimental design (QE)

SOURCE: SEI/CEEP analyses of IES data.

Table 2. Brief summary of main outcomes for each of the eight program evaluations that have published at least preliminary results

Program evaluations	Evidence of effect
<p>Evaluation of the DC Opportunity Scholarship program impacts after 2 years</p>	<p>Positive Effect: The Program had a positive impact on overall parent satisfaction and parent perceptions of school safety. Among the secondary analyses of subgroups, there were impacts on math for students who applied from non-SINI schools and for those with relatively higher pre-Program test scores. However, these achievement outcomes may be a mere by-product of multiple analyses.</p> <p>Adverse Effect: N/A</p> <p>No Effect: The Program had no effect on students' reports of satisfaction and safety. There were no significant impacts on reading achievement or math achievement from the offer of a scholarship or from the use of a scholarship. Students who were offered a scholarship reported similar levels of dangerous activities at school compared to those in the control group; there was also no impact on student reports of school safety from using a scholarship. Overall, there were no impacts of the OSP from being offered or using a scholarship on students' satisfaction with schools.</p>
<p>Evaluation of Enhanced Academic Instruction in Afterschool programs: First year impact findings</p>	<p>Positive Effect: The enhanced program provided students with 30% more hours of math instruction over the school year, compared with students in the regular afterschool program group. There are significant impacts for the enhanced math program on student achievement, representing 8.5% more growth over the school year for students in the enhanced program group as measured by the SAT 10 total math score.</p> <p>The enhanced program provided students with 20% more hours of reading instruction over the school year, compared with students in the regular afterschool program group. There are positive and statistically significant program impacts on one of the two measures in the DIBELS fluency test (reading measure).</p> <p>Adverse Effect: N/A</p> <p>No Effect: Neither the math nor English programs produced significant impacts on any of the three school-day academic behavior measures: student engagement, behavior, or homework completion.</p>
<p>National Evaluation of Early Reading First: Final Report to Congress</p>	<p>Positive Effect: ERF increased the number of hours of professional development that focused on language and early literacy topics. ERF improved the quality of assistant teachers' interactions with children; organization of the classroom environment; lesson planning; quality of the classroom-learning environment; oral language use by both the lead and assistant teachers; book-reading practices that include introducing new vocabulary, using expressive voice, and asking open-ended questions, and improved phonological awareness activities and print and letter knowledge materials</p> <p>ERF had a statistically significant positive effect on children's print and letter knowledge.</p> <p>Adverse Effect: Despite earlier concerns, ERF did <i>not</i> affect children's social-emotional skills.</p> <p>No Effect: ERF had no statistically discernable impact on children's phonological awareness or oral language.</p>
<p>Reading First Impact Study Interim Report</p>	<p>Positive Effect: Reading First increased instructional time spent on the five essential components of reading instruction promoted by the program (phonemic awareness, phonics, vocabulary, fluency, and comprehension). Study sites that received their Reading First grants later in the federal funding process (between January and August 2004) experienced positive and statistically significant impacts both on the time first and second grade teachers spent on the five essential components of reading instruction and on first and second grade reading comprehension. Reading First increased highly explicit instruction in grades one and two and increased "high quality student practice" in grade two.</p> <p>Adverse Effect: N/A</p> <p>No Effect: On average, across the 18 study sites, Reading First did not have statistically significant impacts on student reading comprehension test scores in grades 1-3.</p>
<p>The Enhanced Readings Opportunity: Early Impact and Implementation Findings</p>	<p>Positive Effect: ERO programs produced an increase of 0.9 standard score point on the GRADE reading comprehension subtests.</p> <p>Adverse Effect: N/A</p> <p>No Effect: N/A</p>

<p>Evaluation of Title 1: Final Report</p>	<p>Positive Effect: For the third-grade cohort, the four interventions combined had positive impacts on phonemic decoding, word reading accuracy and fluency, and reading comprehension. For the fifth-grade cohort, the four interventions combined improved phonemic decoding on one measure. The three word-level interventions combined had similar impacts to those for all four interventions combined. There were impacts on both measures of phonemic decoding for students in the fifth-grade cohort. For students in the third-grade cohort, Failure Free Reading (the only word level plus comprehension program) had an impact on one measure of phonemic decoding, two of the three measures of word reading accuracy and fluency, and one measure of comprehension. Being in one of the interventions reduced the reading gap in Word Attack skills by about two-thirds for students in the third-grade cohort.</p> <p>Adverse Effect: The four interventions combined led to a small reduction in oral reading fluency.</p> <p>No Effect: For the third-grade cohort, impacts were not detected for all measures of accuracy and fluency or comprehension. The three word-level interventions combined did not have an impact on either measure of comprehension for students in the third grade cohort. Failure Free Reading did not have any impacts for students in the fifth-grade cohort. The interventions did not improve PSSA scores.</p>
<p>Effectiveness of Reading and Mathematics Software Products</p>	<p>Positive Effect: N/A</p> <p>Adverse Effect: N/A</p> <p>No Effect: Test scores were not higher in classrooms using the selected reading and mathematics software products.</p>
<p>Third National Even Start Evaluation</p>	<p>Positive Effect: N/A</p> <p>Adverse Effect: N/A</p> <p>No Effect: Analysis of pretest compared with posttest data did not show that Even Start children and adults performed better than control group children and adults (see St.Pierre, Ricciuti, Tao, et al, 2003)</p>
<p>National Evaluation of the 21st Century Community Learning Centers Program</p>	<p>Positive Effect: Treatment-group students reported feeling safer afterschool than control-group students.</p> <p>Adverse Effect: Treatment-group students were more likely than control-group students to be with adults who were not their parents and less likely to be with their parents afterschool. Teachers reported lower levels of effort and achievement for treatment-group students relative to control-group students. Treatment-group students were more likely than control-group students to engage in negative behaviors during the school day.</p> <p>No Effect: There was no impact of the program on the frequency of self-care. Treatment-group students scored no better on reading tests than control-group students and had similar grades in English, mathematics, science, and social studies. There also were no differences in time spent on homework, preparation for class, and absenteeism. There was no impact of the program on parental involvement in school.</p>

Consistent data collection means data are collected in the same way and at the same time from the intervention and control groups. All of the studies in table 1 described data collection procedures in detail that were consistent across treatment and control groups. Seven of the eight studies had at least one follow-up data collection point in the reporting of their results.

Each of the eight studies provided a detailed description and justification of the outcome measures used in the study and of the analytic methods used to estimate impact. Meaningful outcomes are those that are of policy or practical importance. If statistically significant results are found in favor of the intervention, it could then be used to improve student and/or teacher performance in areas that really matter. The discussions of outcomes and their possible effect size in the design reports indicates that if the studies

produce significant results, the interventions will, in general, be useful in other schools, districts, etc. Six of the eight studies reported positive and/or negative outcomes, or both positive and negative outcomes. Two studies indicated no impact.

External Reviews of Quality

GPRAs indicator data related to the quality of educational research funded by the Office of Educational Research and Improvement (OERI) and subsequently by IES were available starting with FY 01. Two quite different methods of assessing the quality of research were used in this time period using external review panels.

For FY 01 through FY 04, the GPRAs indicator was:

“The percentage of new research and evaluation projects funded by the Department that are deemed to be of high-quality by an independent review panel of qualified scientists.”

The methodology used for determining the quality of funded projects consisted each year of randomly selecting 20 proposals to be reviewed by a panel of 10 senior scientist expert reviewers. The external reviewers consisted of eminent senior scientists who are distinguished professors, editors of premier research journals, and leading researchers in education and special education. The instructions provided to reviewers on each score sheet asked the reviewers to rate the overall quality of the proposed research, taking into consideration the (a) significance of the project, (b) quality of the project design, (c) qualifications of the personnel, and (d) adequacy of resources. Two reviewers rated each proposal using a 9-point Likert-type scale where 1 represented “very poor quality”, 3 represented “poor quality”, 5 represented “good quality”, 7 represented “high quality” and 9 represented “very high quality”; and a mean score was calculated for each project using the ratings of both reviewers. High quality was defined as receiving a mean rating of 6.5 or higher on this 1-9 scale.

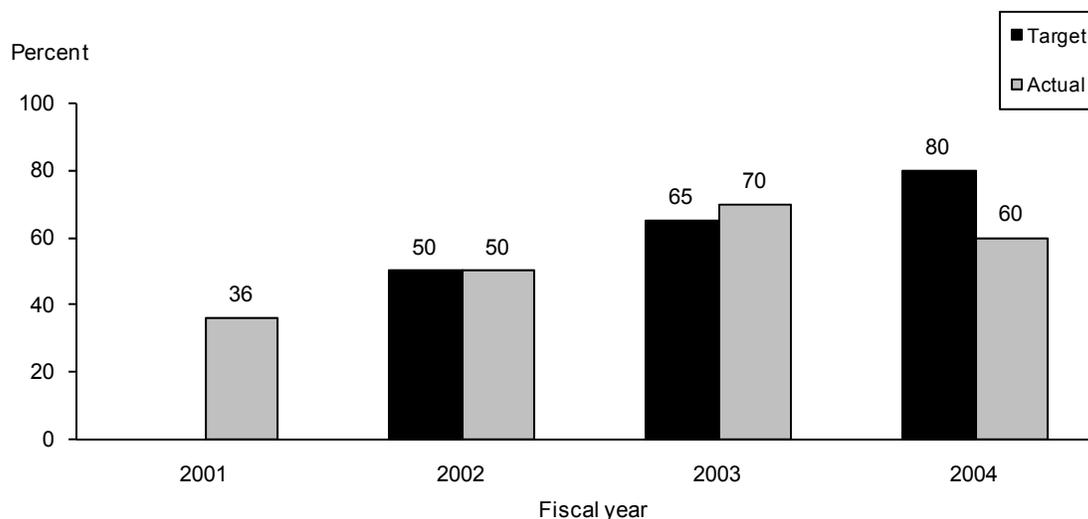
This external and independent review process of proposed projects to determine quality was discontinued after FY 04. Beginning in FY 05, the data collection procedure for this GPRAs indicator was changed in a manner that changed the definition and interpretation of the data. Rather than randomly select proposed funded projects to send to an external review panel that was independent from the funding decisionmaking process, the GPRAs indicator was changed to utilize the review scores that were part of the newly established scientific peer-review process. The scientific peer-review panels were comprised of 12 to 20 leading researchers, and the overall panel scores (scale of 1 to 5, with 1 being outstanding and 5 being poor) were used to determine the percentage of funded projects deemed to be of high quality.

Rigor

Those projects with an average review panel score of 2 or less were considered to be of high quality for the purposes of this indicator.

Figure 2 provides a graphical representation of the GPRA indicator data from FY 01 through FY 04. As noted in the figure below, and taking into consideration the slightly skewed results for FY 04⁸, the reviews conducted by the external panel of senior scientist expert reviewers indicates a steady increase in the quality of proposed projects. Whereas only 36 percent of proposed projects were deemed to be of high quality in FY 01 under OERI, after the establishment of IES, almost twice as many (approximately 70%) were rated as high quality in FY 03 and FY 04 (i.e., with the correction for the extreme outlier). Even with the adjustments for the extreme outlier reviewer in FY 04, however, the actual percentage of proposed projects deemed to be of high quality fell short of the targeted goal of 80 percent. Unfortunately, there are no data available related to the consistency of ratings across the funding years. For example, it is difficult to assess the degree to which external reviewers in later years simply scored more generously than reviewers from previous years. Including some measure of reliability of these measures of quality would increase the meaningfulness of these findings in terms of changes over time.

Figure 2. Percentage of IES funded new research and evaluation projects that are deemed to be of high quality by external review panel: 2001-2004

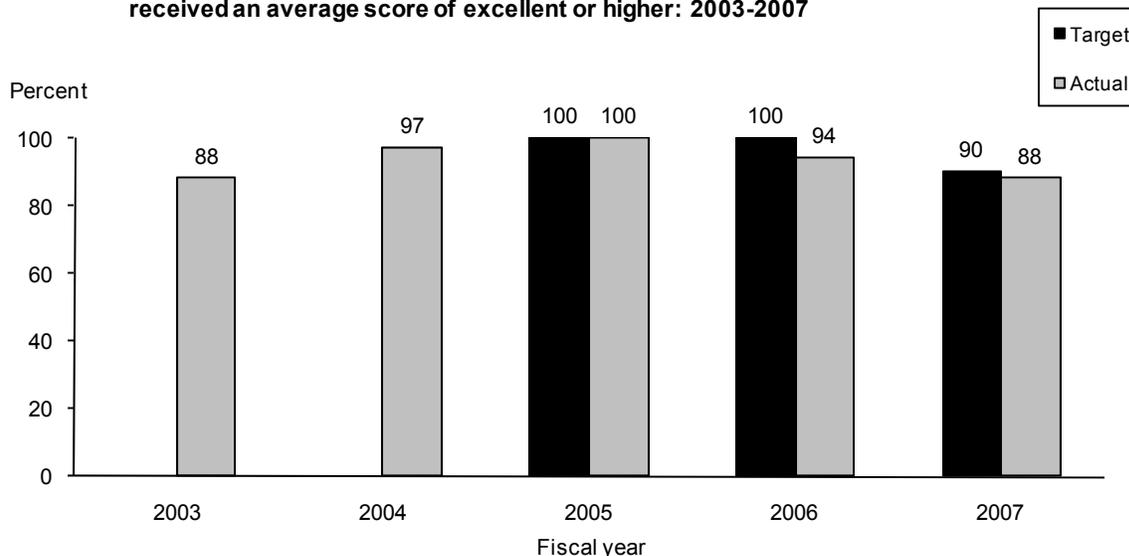


SOURCE: U.S. Department of Education, FY 2004 Program Performance Report.

Explanatory notes on the indicators state that “in 2004, the scores of one reviewer were extreme outliers – greater than 3.8 standard deviations below the average ratings of the other reviewers. If these scores were removed, the percentage of new projects deemed to be of high quality would be 70 percent.”

Figure 3 reflects the new rating process instituted in 2005, and shows the targeted percentages (for FY 05 to FY 07) and actual percentages of new research proposals funded by NCER that received an average score of excellent or higher using the funding panel review process. As noted in the figure, although slightly below targeted percentages for the last 2 fiscal years, the overall percentage of funded NCER proposals that were rated excellent or higher by an independent review panel of qualified scientists has been consistently high (88% to 100%) since the inception of the new scientific peer-review process.

Figure 3. Percentage of targeted and actual new research proposals funded by NCER that received an average score of excellent or higher: 2003-2007



SOURCE: U.S. Department of Education, FY 2007 Program Performance Report.

Discrepancies in these two methods of calculating quality are evident in examining the available data for FY 03 and FY 04. For these two fiscal years, the new scientific review process was already implemented, but the original GPRA indicator related to quality was still being used. Therefore, for these two fiscal years, both types of data are available and provide the basis for a comparison. For FY 03, 70 percent of the randomly selected funded proposals were rated high quality by the external reviewers who were independent from the funding decision process; however, 88 percent of the funded proposals received ratings of excellent or higher as part of the funding review scoring process. Similarly, for FY 04, 60 percent to 70 percent (depending on the inclusion of the extreme outlier) of the randomly selected funded proposals were rated high quality by the external reviewers who were independent from the funding decision process; however, 97 percent of the funded proposals received ratings of excellent or higher as part of the funding review scoring process.

As opposed to the process described earlier of using an external review team not associated in any way with funding decisions (i.e., prior to FY 04), the current GPRA indicator is based on the rating scale used

Rigor

for the actual funding decisions. Although the reviewers for the current GPRA indicator are external to, and independent of, IES and the U.S. Department of Education, these reviewers are not independent of the funding decisionmaking process. Using an indicator that is based on the actual panel scoring process itself necessitates a very high percentage of proposals being rated high quality—as typically studies would not be funded unless they met the cut-off value. More explicitly, the overall scores of these panels are used to make decisions about which proposals to fund, with proposals rated 2.0 or less generally being funded. Therefore, the percentage of funded proposals with a score of 2.0 or less should be close to 100 percent by definition of the funding process itself. Therefore, the most current iteration of the GPRA indicator related to quality does not appear to offer meaningful data related to the quality or rigor of NCER-funded proposals.

Rigor of Evidence from IES-Supported Efficacy, Effectiveness and Research Projects

To what extent are these research and evaluation studies producing (or likely to produce) valid evidence?

Examining the degree to which IES funded rigorous studies and evaluations of educational interventions is one way to examine whether or not the level of rigor is increasing over time. Optimally, however, such a review would also go beyond looking at proposals to examine the degree to which IES funded studies have been able to produce valid and rigorous evidence using the WWC standards of evidence. Even more compelling might be the ability to compare the efficacy and effectiveness research and evaluations funded in the OERI era to that funded under the IES with respect to these criteria. We found this review process to be quite difficult to undertake for several reasons.

First, performing research which produces evidence of efficacy and effectiveness takes time, and reports and published findings from grant-funded research studies were difficult to find. For the purposes of this evaluation, IES provided final reports from 2002 Preschool Curriculum Evaluation Projects funded under IES, but not for any other initiatives during that year, nor any reports after 2002. Review of annual performance reports and of published articles yielded considerable variation in describing the actual study implementations and/or current findings, such that consistent review across studies was not feasible. Therefore, for our data examining the degree to which grant-funded research findings met WWC evidence standards, we were limited to the 2002 Preschool Curriculum studies (N = 12). Second, we were only able to obtain a small, nonrandom sample of final reports from OERI-funded research projects that addressed causal questions (N = 12 from 1999–2000). Because the small sample of reports does not necessarily represent the domain of research conducted over the 11-year period of interest, we do not report any data on these studies across time.

Therefore, this evaluation was limited in its ability to address the extent to which IES research and evaluation studies are producing (or likely to produce) valid evidence. The data for this evaluation question consist of three primary sources: (1) PART data related to the number of IES-supported interventions meeting WWC standards of evidence of effectiveness, (2) stakeholder perceptions of rigor of research funded by IES, and (3) quantity of peer-reviewed publications from IES/OERI funded grants. Each of these data sources is discussed below.

Number of IES-supported interventions meeting WWC standards of evidence of effectiveness

The WWC provides educators, policymakers, researchers, and the public with a central source of scientific evidence of what works in education through high-quality reviews of programs, products, practices, and policies intended to improve student outcomes. For a study to be eligible for review by the WWC it must be a RCT or an appropriate quasi-experiment (e.g., groups created by equating on pretest or prior measure; regression discontinuity designs, or single case studies). RCTs are defined as studies where the assignment of participants to treatment and control is functionally haphazard or truly random. The studies submitted for review for WWC undergo three stages of review.

Stage 1 is a review to determine the relevance of the study and sample, and the appropriateness of the data collected. In particular, studies that are related to the topic specified in the competition and have an educationally relevant sample of participants meet the first criteria. Studies that utilize outcome measures related to academic or teaching success and that report adequate information (e.g., psychometric and descriptive) about the measures meet the second criteria.

Stage 2 is a review conducted to assess the degree to which the study provides evidence for efficacy or effectiveness that meets the evidence standards set out by the WWC. In particular, the review focuses on the following dimensions:

Type of study design. Well-implemented RCTs are assigned Meets Evidence Standards without Reservations and well-implemented quasi-experimental studies are assigned Meets Evidence Standards with Reservations designations;

Reportable and reasonable effect size—preferably in standardized mean difference metric;

Evidence of baseline equivalence on the outcome measures for treatment and control groups at onset of study or appropriate adjustment for lack of equivalence during analysis;

Low overall attrition and lack of differential attrition in the treatment and control groups;

Lack of confounds, including lack of intervention contamination (through local history events) and lack of teacher-intervention confounds;

Rigor

Match between randomization and level of analysis—i.e., mismatch potentially overestimates statistical significance.

Stage 3 of the review process is conducted with all studies that Meet Evidence Standards with and without Reservations to describe the groups and settings for which the study has validity. In our analysis, we focus only on the criteria for Stage 2 in examining the rigor of completed research and the potential validity of funded research (in particular—in FYs 2005, 2006, and 2007).

The IES's Program Assessment Rating Tool (PART) includes three separate annual measures based on the number of IES-supported interventions that meet the WWC standards for evidence of effectiveness. These three content areas include interventions with evidence of efficacy in improving student outcomes in reading or writing, in mathematics and science, and in enhancing teacher characteristics with demonstrated positive effects on student outcomes. The data for these PART annual measures are based on WWC principal investigator reviews of initial findings on interventions from IES research grants, such as findings that will have been presented as papers at a convention or working papers provided to IES by its grantees. The WWC principal investigators rate these findings from IES research grants using the WWC published standards to determine whether the evidence from these research grants meets evidence standards of the WWC and demonstrates a statistically significant positive effect in improving achievement outcomes in each of the three content areas.

Table 3 on the next page shows the targeted numbers of interventions and actual interventions for each of the respective PART annual measures. As seen in table 3, IES is currently meeting its targeted goals in interventions demonstrating positive effects in reading and writing and enhancing teacher characteristics, and exceeding targeted numbers of interventions in mathematics and science interventions. In addition, the number of interventions increased between 2006 and 2007 for each of the three content areas.

Table 3. Number of IES-supported interventions meeting What Works Clearinghouse (WWC) standards of evidence of effectiveness: 2005–12

Year	Reading and writing		Mathematics and science		Enhancing teacher characteristics	
	Target	Actual	Target	Actual	Target	Actual
2005	—	1	—	—	—	—
2006	—	3	—	1	—	1
2007	6	6	3	4	3	3
2008	11	—	7	—	5	—
2009	13	—	10	—	7	—
2010	15	—	12	—	10	—
2011	17	—	15	—	12	—
2012	—	—	18	—	15	—

— Not available.

SOURCE: U.S. Department of Education, FY 07 Program Performance Report.

In addition to the PART annual measures, long-term outcome measures related to these three content areas are included in PART. These three measures are as follows:

The minimum number of IES-supported interventions on reading or writing that are reported by the WWC to be effective at improving student outcomes by 2013–2014. Target: 15.

The minimum number of IES-supported interventions on mathematics or science education that are reported by the WWC to be effective at improving student outcomes by 2013–2014. Target: 12.

The minimum number of IES-supported interventions on teacher quality that are reported by the WWC to be effective at enhancing teacher characteristics with demonstrated positive effects on student outcomes by 2013–2014. Target: 10.

Given the annual measures noted in the table above for each of these content areas, it is surprising that the established targets for 2013–2014 are lower than the established annual targets for reading and writing (i.e., annual targets 17 by 2011 whereas long-term targets 15 by 2013–2014), mathematics and science (i.e., annual targets 18 by 2012 whereas long-term targets 12 by 2013–2014), and enhancing teacher characteristics (i.e., annual targets 15 by 2012 whereas long-term targets 10 by 2013–2014). IES has stated that these differences are due to the long-term targets referring to interventions that show positive impacts when implemented at scale, whereas for the annual targets the evaluations do not need to be at scale. However, this distinction is not clear from the publicly available PART data. Although the explanatory notes state the long-term goals are interventions “that are effective and can be widely deployed,” it is not clear from these notes that the interventions need to already show impact when implemented at scale to be included in the PART long-term assessment data.

Rigor

Stakeholder Perceptions of Increased Rigor of Research Funded by IES

As noted previously, SEI/CEEP conducted key stakeholder interviews to supplement extant data. Interviews were completed with key stakeholders from the following organizations and associations: American Educational Research Association (AERA), American Psychological Association (APA), National Academy of Sciences, Council of the Great City Schools, Knowledge Alliance, and National Sorority of Phi Delta Kappa. Interpretation is somewhat limited due to the small numbers of education-related organizations represented. However, the stakeholders included in the data represent some of the largest and most representative education-related organizations in the nation (e.g., AERA and APA); and the interview responses represent these persons' perceptions of the views and opinions of their broader constituencies, rather than the individual opinions of six persons. Therefore, the data from these six interviews do provide some valuable insight into perceptions of IES impact, particularly when interpreted within the context of other available data.

All stakeholders interviewed strongly believed that IES had increased the quality of research being conducted within the field of education. All interviewees also believed that the emphasis on rigor is significantly more pronounced within IES than it was during the era of OERI. One stakeholder referred to OERI as having a "soft edge" whereas IES has a "hard edge" with a focus on outcomes and impact. Representative comments related to the perceived impact of IES on rigor in the field of education included the following:

"I give IES good grades for increasing rigor. They grabbed the research community by the lapels and forced them to be more conscious of quality. The education community needed and continues to need a kick in the pants regarding standards of scientific inquiry, and that is what they got from IES."

"IES has definitely pushed the field to be more rigorous. Some claim they have gone too far, but I am not sure that's correct. I am happy to see them overcorrect. The field has not been good at pushing the field forward."

"Russ has made impressive changes. The attention to wanting to fund high quality research and articulate the importance of rigor and look at the review process has put IES in a much stronger and more credible position ... He has elevated the status of IES because of his focus on rigor."

The strong consensus on the positive impact of IES on increasing the rigor of education research does not imply that these stakeholders did not have any concerns with the strong emphasis on RCTs. But, as stated by one person, "the religious belief in RCTs carries baggage, but it has put IES on the map in terms of rigor." Several stakeholders noted the negative impacts of the strong focus on rigor, and particularly the focus on RCTs. Comments related to these negative impacts included the following:

“The overemphasis on RCTs led to the perception that other methodologies that can produce evidence are no longer tolerated or only tolerated at best as 2nd or 3rd best, and only when the gold standard is not feasible. IES has suffered because of its lack of respect for other methods. It has had the unintended consequence of narrowing the field.”

“Even if IES empirically funds the full spectrum, but highlights, showcases and articulates only the small band of RCTs – opportunities are lost.”

Several stakeholders noted that the position of IES related to rigorous research and RCTs has been modified over time, and is now more inclusive of other methodologies. These stakeholders noted the perceived shift was a positive step in still maintaining rigor, and RCTs as appropriate based on the research question, without excluding other valid and rigorous methodologies.

Quantity of Peer-Reviewed Publications from IES/OERI Funded Grants

The number of peer-reviewed publications from IES and OERI grants was obtained through searches of JSTOR, ERIC, and the Indiana University Library system, along with listings of publications on IES project abstract pages and professional pages of IES/OERI principal investigators. A publication was counted if it met the following criteria: (1) published in peer-reviewed journal; (2) published within a reasonable time frame (e.g., after the date of the grant); (3) referenced IES funding; and (4) appeared to report content and findings relevant to the grant project.⁹

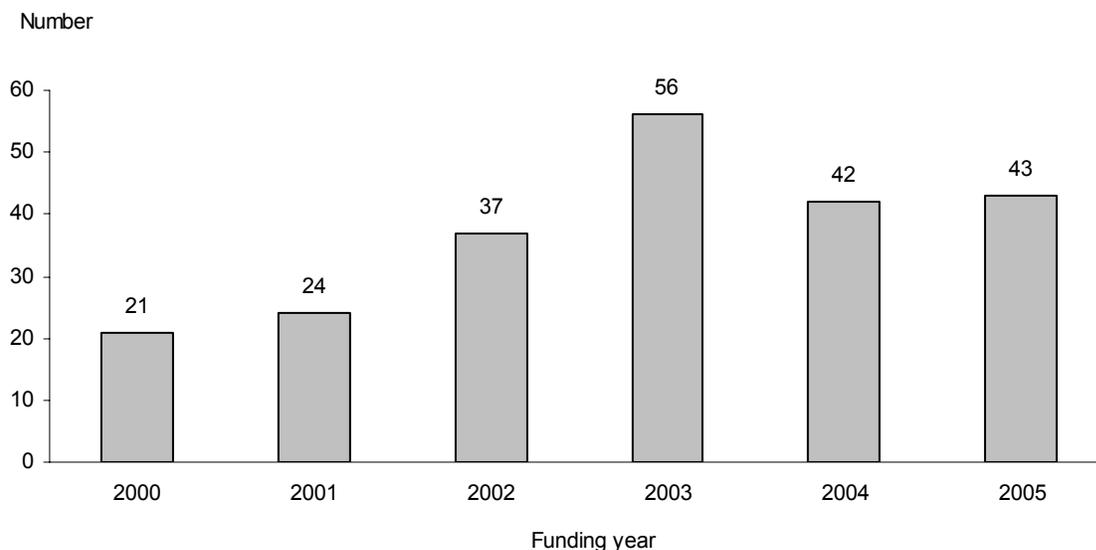
Figure 4 depicts the number of peer-reviewed publications for each of the OERI and IES funded grants from 2000 through 2005, with the numbers of publications noted by funding year as opposed to the actual year of publication. In other words, regardless of the actual year of publication, the journal article is included within the grant year the funding was first received. For example, a 2005 journal article resulting from a 2001-funded grant would be indicated in the graph as a peer-reviewed publication for funding year 2001. From figure 4 on the next page, it appears that more peer-reviewed publications were published from grants funded during the first 2 years of IES than during the grants funded during the last 2 years of OERI (93 versus 45). For the 2 years of OERI data (2001 and 2002) versus the 4 years of IES

⁹ It should be noted that although the publication search was systematic and wide-reaching, it may not be comprehensive in nature, nor may all of the publications counted be directly related to one specific grant, as many publications referenced multiple federal or other funding sources, or did not reference a specific funding source at all. Grant programs reviewed under IES were funded through NCER and did not include NAEP Secondary Data Analysis, Small Business Innovation Research, or National Research and Development Center grants.

Rigor

data (2002 through 2005) this translates to an average of 11.3 peer-reviewed publications per year for OERI grants, and an average of 44.5 peer-reviewed publications per year for IES grants.¹⁰

Figure 4. Number of peer reviewed publications referencing OERI (2000-2001) and IES (2002-2005) funded grants



SOURCE: SEI/CEEP analyses of IES data.

The Potential to Produce Valid and Rigorous Evidence of Effectiveness

To what extent are these research and evaluation studies producing (or likely to produce) valid evidence?

To provide data on the *potential* of IES grant-funded projects to produce rigorous and valid findings, we reviewed the funded proposals for FYs 2004-2007. For these proposals, we were unable to determine if certain criteria had been met since the research had yet to be completed; however, we were able to code whether or not the proposal itself dealt with particular evidence criteria in the proposal narrative and provided solutions for difficulties that might arise. To the degree that the study designs took into account various elements of rigorous research, they are more likely to produce valid and rigorous findings. Funded proposals were coded for FYs 2004 through 2007¹¹ on ten dimensions of high quality research designs.¹² Three criteria dealt with prestudy planning:

¹⁰ To an unknown extent, publications are admittedly a lagging indicator of field-initiated research projects.

¹¹ These years were chosen because little published or disseminated findings are currently available—therefore it makes sense to assess their potential to produce rigorous and valid research.

Conducting strong power analyses;
Having an adequate sample size to detect a meaningful effect; and
Choosing the appropriate level and method of randomization (or matching, in the case of quasi-experimental designs).

The next three criteria were related to the quality of the outcomes and the data collection procedures:

Outcomes were deemed valid and appropriate;
Systematic data collection procedures were used with both treatment and control groups, and
Longitudinal or follow-up measures were collected at least 1 year after the intervention.

Finally, four criteria were coded dealing with the actual implementation of the study itself, and potential threats to validity and appropriate analysis. These indicators included:

Consideration of baseline equivalence of treatment and control groups;
Appropriate analytic methods;
Appropriate consideration of attrition issues, including intent-to-treat analyses and minimizing effects,
and
Consideration of and solutions to other internal and external validity threats (including teacher-intervention confounds, history and development threats, generalizability issues, etc.).

The coding process involved assigning a Met or Not Met designation for each of the ten characteristics above by examining reviewer comments. If a reviewer pointed out a particular characteristic as a strength of the study or did not mention it as a weakness, the criteria was designated as Met. If the reviewers designated a particular characteristic as a potential weakness, it was marked as Not Met. The benefit of this particular coding scheme is that it is based on the expert panel reviews that had already been conducted on the funded proposals, and does not add another, potentially conflicting, layer of review.

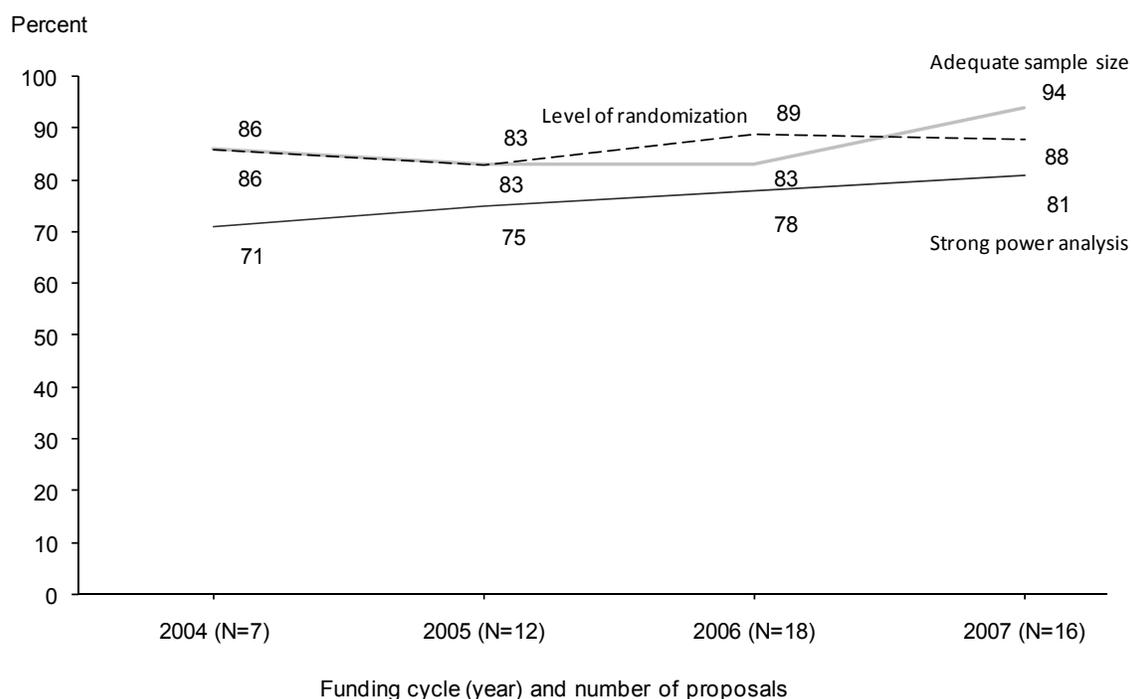
¹² The 10 criteria used to code the proposals were adapted from the WWC Evidence standards and from the descriptions of the RFAs for the grant proposals themselves. The main difference between these and WWC criteria is that these are prospective considerations of proposals; thus “evidence of effects” is not applicable, and a category for “appropriate analytic methods” has been added instead.

Rigor

The figures below show the percentage of funded proposals in FY 04, FY 05, FY 06, and FY 07 that met each of the criteria.

The first figure, figure 5 below, shows the ratings for the prestudy criteria of correct power analysis, adequate sample size, and appropriate level of randomization. From this figure, one can see that the use of correct power analysis techniques has increased from 71 percent to 81 percent over the 4 years, while the percentage of funded research studies addressing causal questions having an adequate sample size and using the appropriate level and method of randomization has increased from 86 percent to 88 percent, and 86 percent to 94 percent, respectively—indicating that a high percentage of funded studies have these characteristics.

Figure 5. Percentage of proposed studies addressing pre-study criteria of correct power analysis, adequate sample size, and appropriate level of randomization: 2004-2007

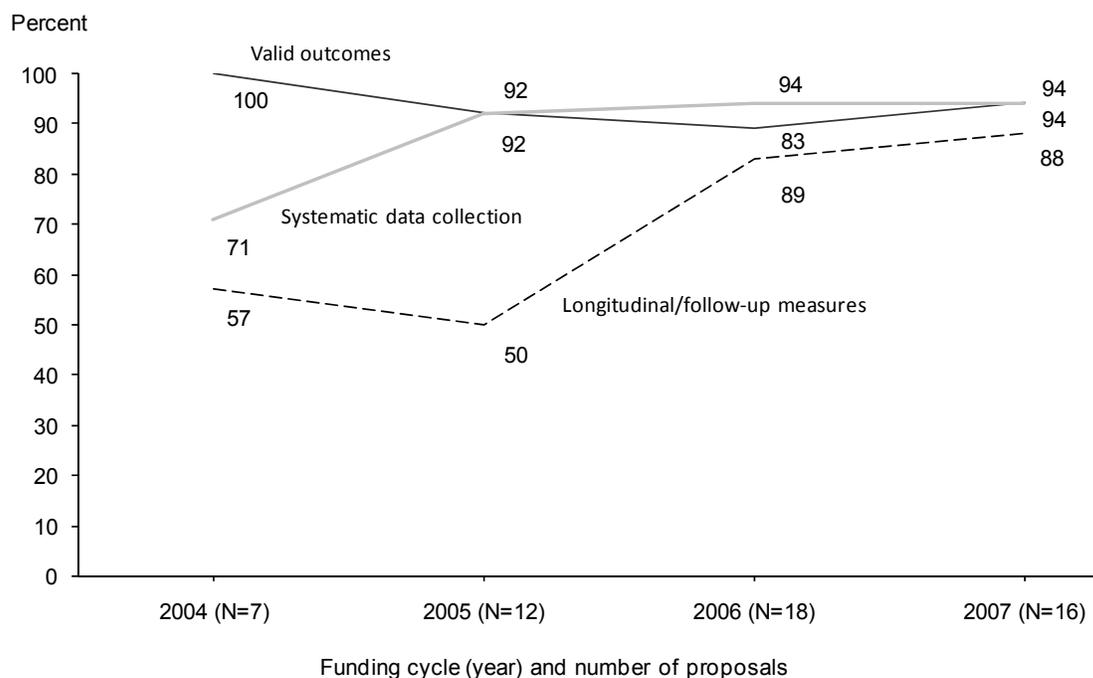


SOURCE: SEI/CEEP analyses of IES data.

The second figure, figure 6, shows the ratings for the quality of the outcomes and the data collection procedures criteria of valid outcomes, systematic data collection, and longitudinal or follow-up measures at least a year after the intervention. From this figure, one can see that the use of valid outcomes has maintained at a high level over the 4 years, while the percentage of funded research studies addressing causal questions using systematic data collection procedures for both treatment and control groups has increased 71 percent to 94 percent, and the percentage of studies using longitudinal methods has increased

from 57 percent to 88 percent, respectively—indicating that a high percentage of funded studies have these characteristics in 2007.

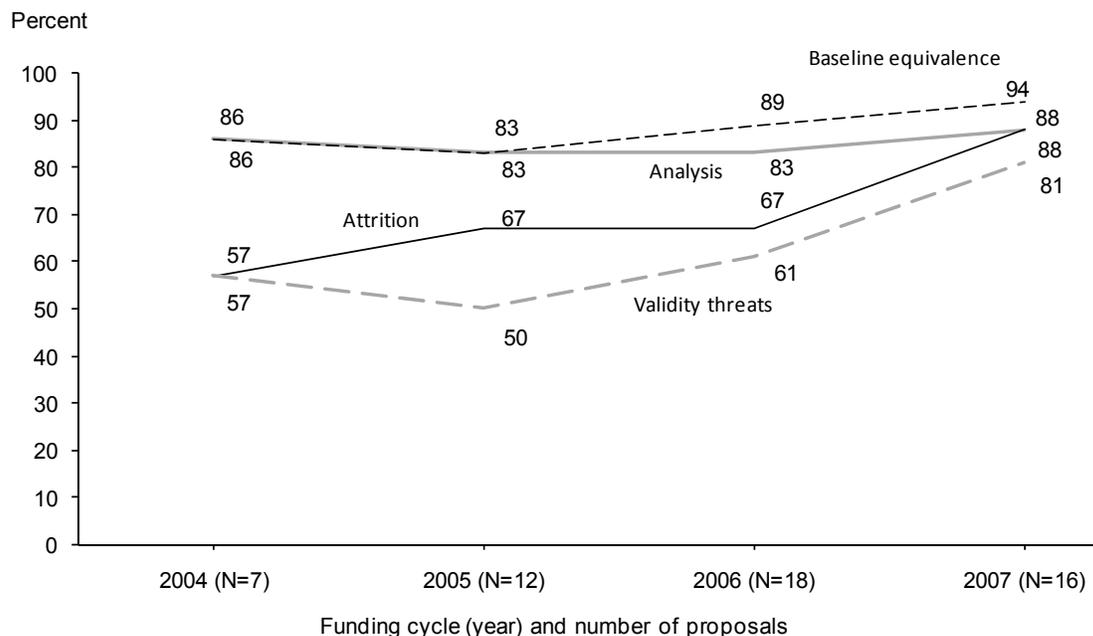
Figure 6. Percentage of proposed studies with valid outcomes, systematic data collection, and longitudinal or follow-up measures: 2004-2007



SOURCE: SEI/CEEP analyses of IES data.

The third figure shows the ratings for how well proposals addressed implementation issues, including potential threats to validity and appropriate analysis criteria of baseline equivalence, appropriate analytic methods, consideration of attrition issues, and other threats to study validity. From this figure, using appropriate analytic methods and properly considering issues of baseline group equivalence has maintained at a high level over the 4 years, while the percentage of funded research studies addressing causal questions that appropriately address attrition issues and other threats to study validity has increased from 57 percent to 88 percent, and 57 percent to 81 percent, respectively—indicating that a high percentage of funded studies have these characteristics in 2007.

Figure 7. Percentage of proposed studies addressing analysis, attrition, validity threats and baseline equivalence: 2004-2007



SOURCE: SEI/CEEP analyses of IES data.

These combined results provide evidence that the studies being funded by IES have a high potential for generating rigorous and valid evidence of effectiveness, if the study parameters proposed can be maintained during the study itself, or modified in rigorous ways if necessary.

Capacity of the Field to Conduct Rigorous Education Research

Discussions during the June 2002 hearings on the reauthorization of the Office of Education Research and Improvement (OERI)¹³ included statements regarding the need to increase capacity in the education research community. Subsequently, the mission of the Institute of Education Sciences (IES) included not only a focus on increasing rigor within the education community, but also a parallel focus on increasing the capacity of the field to conduct rigorous research. The primary mechanisms used by IES for increasing the capacity of the field to conduct rigorous education research are predoctoral and

¹³ Reauthorization of the Office of Education Research and Improvement (OERI): Hearing of the Committee on Health, Education, Labor, and Pensions, United State Senate. One Hundred Seventh Congress Second Session on Examining proposed legislation authorizing funds for the office of education research and improvement, department of education, focusing on organizational structure, budget and technical assistance systems. June 25, 2002.

postdoctoral fellowships. As noted on the IES website¹⁴ the purpose of the predoctoral program is “to address the shortage of education scientists who are prepared to conduct rigorous education research... [and to] support the development of a new generation of education scientists.” Similarly, the purpose of the postdoctoral program is “to increase the supply of scientists and researchers in education who are prepared to conduct rigorous evaluation studies, develop new products and approaches that are grounded in a science of learning, design valid tests and measures, and explore data with sophisticated statistical methods.”¹⁵ Therefore, within the scope of the available data and resources, this evaluation addresses the following questions related to these two fellowship programs: To what extent has IES increased the number and quality of pre- and postdoctoral scientists? To what extent are pre- and postdoctoral scientists funded through IES programs likely to contribute to the quantity and quality of rigorous evidence related to education practice?

In addition, this section also provides data related to two other mechanisms that have been used by IES to increase the capacity of the field to conduct rigorous education research. First, available extant data related to NCES on the various databases are provided. Second, accessible data related to other IES trainings efforts besides the pre- and postdoctoral fellowships are provided.

NCER Pre- and Postdoctoral Fellowship Programs

To what extent has IES increased the number and quality of pre- and postdoctoral scientists?

Numbers of Pre- and Postdoctoral Scientists. The National Center for Education Research (NCER) supports 15 interdisciplinary predoctoral research training programs. However, five (5) of these awards for predoctoral research training programs were made in July 2008 for FY 08 Second Phase, and therefore are not specifically discussed within the scope of this report given the recent nature of these awards. This report focuses on initial ten predoctoral research trainings funded by IES: the five institutions of higher education receiving funding in 2004, and the five additional institutions receiving funding in 2005.

According to the IES website, these predoctoral students “are being trained to develop education interventions (e.g., curricula, professional development) that are grounded in a science of learning; to evaluate education programs, practices, and policies using rigorous and well-implemented experimental

¹⁴ <http://ies.ed.gov/ncer/projects/program.asp?ProgID=16> (retrieved July 21, 2008).

¹⁵ <http://ies.ed.gov/ncer/projects/program.asp?ProgID=14> (retrieved July 21, 2008).

Rigor

and quasi-experimental designs; and employ sophisticated statistical methods to examine large state and local datasets to identify potential solutions to education problems.”¹⁶ Findings related to the number and quality of predoctoral scientists funded by IES include the following:

A total of 242 predoctoral fellows have been funded from 2004 through 2008 at 10 institutions of higher education.¹⁷

Approximately 16.5 percent of the predoctoral fellows (N=40) from 2004 through 2008 are racial-ethnic minorities. This percentage is lower than national survey statistics from 2005 indicating that approximately 21.4 percent of all doctoral recipients with degrees related education research were racial-ethnic minorities (Hoffer et al., 2006).

Approximately 7.4 percent (18 of 242 total) of the predoctoral students have left the IES fellowship program (e.g., left with masters, transferred to another university, still in doctoral program but dropped from fellowship, or dropped out/left academia).

To date the number of completed Ph.D.s who are employed (including summer 2008 with jobs lined up) is 37, with 24 postdoctoral fellows from 2004 programs, and 13 postdoctoral fellows from 2005 programs. Interestingly, one institution of higher education contributed more than half of the completed Ph.D.s (i.e., 16 of 24) for the 2004 cohort of grantees. However, given that most of the predoctoral fellows have not been participating in the programs long enough to have obtained their Ph.D.s and entered the workforce, at this point in time the numbers of completed Ph.D.s who are employed can only be considered very preliminary data.

In terms of the numbers of postdoctoral scientists, IES partnered with the American Psychological Association in FY 04 to establish new postdoctoral fellowships to provide training opportunities for psychologists in education research. Based on the success of this program NCER subsequently announced a new postdoctoral training grant program open to education scientists in any discipline. Currently, NCER supports seventeen interdisciplinary postdoctoral research training programs across 14 institutions of higher education. Three institutions received two postdoctoral fellowship awards in

¹⁶ Retrieved from <http://ies.ed.gov/ncer/projects/program.asp?ProgID=16> on July 21, 2008.

¹⁷ Note: This total number of predoctoral fellows represents the number of unique individuals funded through this program, as opposed to Table 8 that represents the numbers of predoctoral fellows funded in any given year. Therefore, an individual may appear multiple times in the counts for Table 8.

overlapping years under different principal investigators. Of these 17 awards, 6 were funded in 2005, 4 were funded in 2006, three 3 funded in 2007, and 4 were funded in 2008.

Postdoctoral fellows are generally supported for 2 to 3 years (with a maximum of 4 years), and each institution can request funds for up to four fellows. In addition, five new grants for Postdoctoral Special Education Research Training Fellowships were awarded in July 2008 as part of FY 08 Second Phase funding. However, given that these awards were just announced, they are not included in the findings for this evaluation.

According to the IES website, the fellows involved in the Postdoctoral Research Training Program should: “(a) gain the breadth of skills and understanding necessary to conduct rigorous applied research in education, and (b) develop the capacity to independently carry out such research, including applying for grant funding and submitting results for publication in peer-reviewed journals.”¹⁸ Findings related to the number and quality of postdoctoral scientists funded by IES include the following:

IES has supported a total of 30 postdoctoral fellows between 2005 and 2008 across 17 institutions of higher education.¹⁹

30 percent of the postdoctoral fellows (N=9) from 2005 through 2008 are racial-ethnic minorities. This percentage is higher than national survey statistics from 2005 indicating that approximately 21.4 percent of all doctoral recipients with degrees related education research were racial-ethnic minorities (Hoffer et al., 2006).

One-third (33.3%) of the fellows (N=10) had completed the postdoctoral fellowship and were employed by summer of 2008. However, data were not available related to the specific positions or research agendas of the postdoctoral fellows who have completed the fellowship.

The Program Assessment Rating Tool (PART) also includes benchmark data related to the targeted numbers of pre- and postdoctoral scientists IES hopes to impact through its fellowship training programs. The annual output measure specifically notes “The minimum number of individuals who have been or are being trained in IES-funded research training programs”; and the “explanation” states “The number of

¹⁸ http://ies.ed.gov/ncer/funding/2009_84305B/index.asp?rfa=rprA2 (retrieved July 21, 2008)

¹⁹ This total number of postdoctoral fellows represents the number of unique individuals funded through this program, as opposed to table 4 that represents the numbers of postdoctoral fellows funded in any given year. Therefore, an individual may appear multiple times in the counts for table 4.

Rigor

individuals who receive fellowship support as participants in IES-funded pre- and postdoctoral research training programs will be obtained from grantee reports contained in the official grant files.” Table 4 below notes the established targets for the measure, as well as the actual reported PART data.²⁰ In addition, the table notes the estimated numbers calculated as part of this evaluation for the years not yet publicly reported. Although data indicate the annual output targets were not met for 2007 and 2008, the actual numbers of individuals participating in the IES-funded research training programs was close to the targeted number (92% and 94% of the respective targets).

Table 4. Established targets and actual reported Program Assessment Rating Tool (PART) data and estimates: 2005–12

Year	TOTAL: Target	TOTAL: Actual	Calculated predoctoral	Calculated postdoctoral	Calculated total
2005	—	35	36	0	36
2006	—	97	96	1	97
2007	175	—	143	18	161
2008	230	—	191	26	217
2009	265	—	—	—	—
2010	325	—	—	—	—
2011	400	—	—	—	—
2012	450	—	—	—	—

— Not available.

SOURCE: U.S. Department of Education, FY 07 Program Performance Report.

Quality of Pre- and Postdoctoral Scientists

In terms of the quality of predoctoral students participating in the fellowships, the only available extant data related to quality are GRE scores. Table 5 provides data related to the predoctoral fellows’ average verbal and quantitative GRE scores, as well as the associated percentile ranking range. As noted in the table, the average verbal GRE score was 618 (85th to 89th percentile); and the average quantitative GRE score was 695 (68th to 72nd percentile).

²⁰ These total number of fellows represent the numbers of fellows funded in any given year. Therefore, an individual may appear multiple times in the counts for Table 4. Numbers reported previously represent the number of unique individuals funded through this program.

Table 5. Predoctoral fellow Graduate Record Examination (GRE) scores and percentiles as compared to other groups

Subgroup	Verbal	Percentile ¹	Quantitative	Percentile
Predoctoral fellows	618	85–89	695	68–72
Education ¹	449	43–49	533	31–35
Social sciences ²	487	55–60	563	40–44
All ²	465	49–55	584	44–49
Top education programs ³	543	71–76	618	49–53

¹ Percentile ranking equivalents retrieved from ETS table found at <http://www.ets.org/Media/Tests/GRE/pdf/994994.pdf> Retrieved July 10, 2008.

² Percentile ranking equivalents retrieved from ETS table *General Test Percentage Distribution of Scores Within Intended Broad Graduate Major Field* found at http://www.ets.org/Media/Tests/GRE/pdf/5_01738_table_4.pdf Retrieved July 10, 2008.

³ Walker, G. (2008). Admission requirements for education doctoral programs at top 20 American universities. *College Student Journal*.

SOURCE: U.S. Department of Education, Institute of Education Sciences.

To provide additional context for this data the average GRE scores within intended major fields for education and social science majors are also noted in the table, as well as all seniors and nonenrolled graduates completing the GRE. As indicated in the table, the predoctoral fellows have substantively higher GRE scores for both the verbal and quantitative sections of the GRE, scoring more than 40 percentile points higher than intended education majors on the verbal section and approximately 37 percentile points higher on the quantitative section. Therefore, the students participating in the IES predoctoral training programs appear to be of high quality in terms of applicants to education graduate schools.

These predoctoral fellows also appear to be highly qualified as compared to social science applicants, as well as more generally graduate school applicants. In addition, the table includes a comparison to those doctoral students attending some of the top 20 education programs at institutions of higher education. As compared to successful applicants to the top 20 education doctoral programs in the United States, as identified by Walker (2008), IES predoctoral students score more than 13 percentile points higher on the verbal GRE component and approximately 19 percentile points higher on the quantitative GRE component.

In terms of the quality of postdoctoral students participating in the fellowships, no extant data related to quality were available for the purposes of this evaluation.

NCES Pre- and Postdoctoral Fellowship Programs: Likelihood of Contributing to Quantity and Quality of Rigorous Evidence

To what extent are pre- and postdoctoral scientists funded through IES programs likely to contribute to the quantity and quality of rigorous evidence related to education practice?

Rigor

Within the scope of the available data, the evaluation also examined the likelihood of these funded pre- and postdoctoral fellows contributing to the quantity and quality of rigorous evidence related to education practice. In terms of the predoctoral and postdoctoral training programs, several factors need to be taken into consideration in examining the likelihood of the fellows funded through IES contributing to the quantity and quality of rigorous evidence related to education practice. Assessing likelihood entails examining both the capabilities of the fellows, as well as their willingness and likelihood of conducting rigorous research in the field of education.

In terms of capabilities, within the given data and resources currently available, the following statistics provide some indication of the both the background and experiences of the predoctoral and postdoctoral fellows, as well as their potential to be academically productive:

Refereed Conference Presentations: During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES presenting a total of 662 refereed conference presentations (i.e., 307 for 2006-07 and 355 for 2007-08); and postdoctoral fellows self-reported to IES presenting a total of 132 refereed conference presentations (i.e., 53 for 2006-07 and 79 for 2007-08). This represents an average of 2.7 refereed conference presentations per predoctoral fellow for 2006–08; and an average of 4.4 refereed conference presentations per postdoctoral fellow for 2006–08.

Number Published/In Press Papers. During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES having a total of 126 published/in press papers (excluding conference proceedings). These numbers include 57 published/in press papers for 2006–07 and 69 published/in press papers for 2008. Postdoctoral fellows reported self-reported to IES having a total of 52 published/in press papers (excluding conference proceedings) during the same 2-year period. These numbers include 16 published/in press papers for 2006-07 and 36 published/in press papers for 2008. This represents an average of 0.5 published/in-press papers per predoctoral fellow for 2006–08; and an average of 1.7 published/in-press papers per postdoctoral fellow for 2006–08.

The IES website also notes that “From the Institute's view, a postdoctoral training program would be successful if it produced education researchers who are able to submit competitive applications to the Institute's research competitions.”²¹ To date, one postdoctoral fellow who has completed the training

²¹ http://ies.ed.gov/ncet/funding/2009_84305B/index.asp?rfa=rprA2 (retrieved July 10, 2008).

program has obtained IES funding as a principal investigator or coprincipal investigator. However, given that most fellows have not yet completed training or have only recently completed training, it is still too early to determine success related to this indicator.

The other factor contributing to the likelihood that the pre- and postdoctoral scientists funded through IES are likely to contribute to the quantity and quality of rigorous evidence related to education practice is the postfellowship employment obtained. In terms of predoctoral scientists, given the length of time needed to complete the predoctoral programs, only limited data are currently available related to postemployment obtained by the predoctoral fellows. However, an analysis of the data available for the 35 of 37 completed Ph.D.s who are employed and have available data (including summer 2008 with jobs lined up) provides some preliminary indicators regarding the likelihood of these fellows to contribute to the field of rigorous education research. As also depicted in figure 8, for the 35 completed Ph.D.s with available data:

25.7 percent (N =9) obtained tenure track faculty positions at research universities (Carnegie basic classifications: 5 Research Universities/Very High and 4 Research Universities/High)

14.3 percent (N= 5) obtained tenure track faculty positions at non-research colleges and universities (Carnegie basic classifications: 4 baccalaureate/arts and sciences, and 1 specialized/medical)

22.9 percent (N =8) obtained postdoctoral fellowships, including one with IES (unclear if all education related)

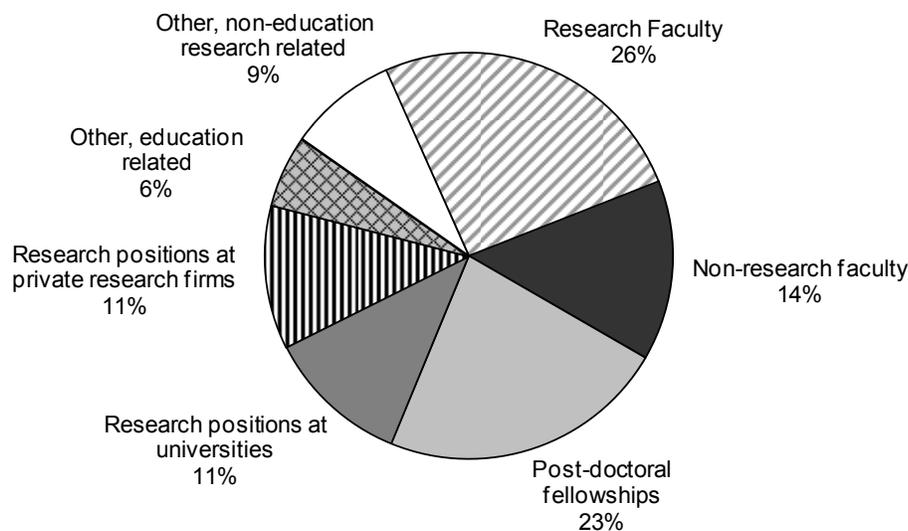
11.4 percent (N=4) obtained research positions at universities

11.4 percent (N=4) obtained research positions at private research firms

5.7 percent (N=2) other education related (State Board of Education, lecturer)

8.6 percent (N=3) Other noneducation research related (CDC, statistician at hospital, project manager at private noneducation firm)

Figure 8. Distribution of current positions of employed IES pre-doctoral fellows who have completed Ph.D. programs: 2008



SOURCE: U.S. Department of Education, Institute of Education Sciences.

In total it appears that approximately 80 percent (N=28) of the employed predoctoral fellows are currently in research positions of some type (i.e., research faculty, postdoctoral fellowships, research position at universities, research position at private research firms, and other noneducation research related).

Although it is possible that faculty at baccalaureate/art and sciences institution of higher education might engage in education research, it is less likely given the nature of these institutions. However, it is not possible to determine from the current data what specific fields of research these individuals will pursue. For example, given the interdisciplinary nature of the predoctoral training programs it is not surprising that many of the faculty positions obtained by the fellows are not within education departments. For example, for those with available departmental data, fields include psychology, social welfare, and economics. What cannot be determined from the available data is whether or not faculty in these noneducation departments will remain engaged in education-related research.

In terms of the ten postdoctoral fellows who have completed their fellowship and were employed by summer 2008, 50 percent (N=5) obtained tenure-track faculty positions at institutions of higher education, 40 percent (N=4) obtained research positions at universities or university research centers, and 10 percent (N=1) obtained a research position at a private research firm.

The Program Assessment Rating Tool (PART) also includes both annual outcome measures and a final outcome measure associated with the postfellowship employment of graduates of IES-supported research training programs. Beginning in 2009, research training programs will be asked to obtain and provide information to IES about the current employment of the individuals who have completed their programs. More specifically, the annual outcome measure establishes benchmarks related to “The minimum number of graduates of IES-supported research training programs who are employed in research positions.” The target is 40 for 2009, with an additional 40 per year resulting in 200 graduates employed in research positions by 2013.

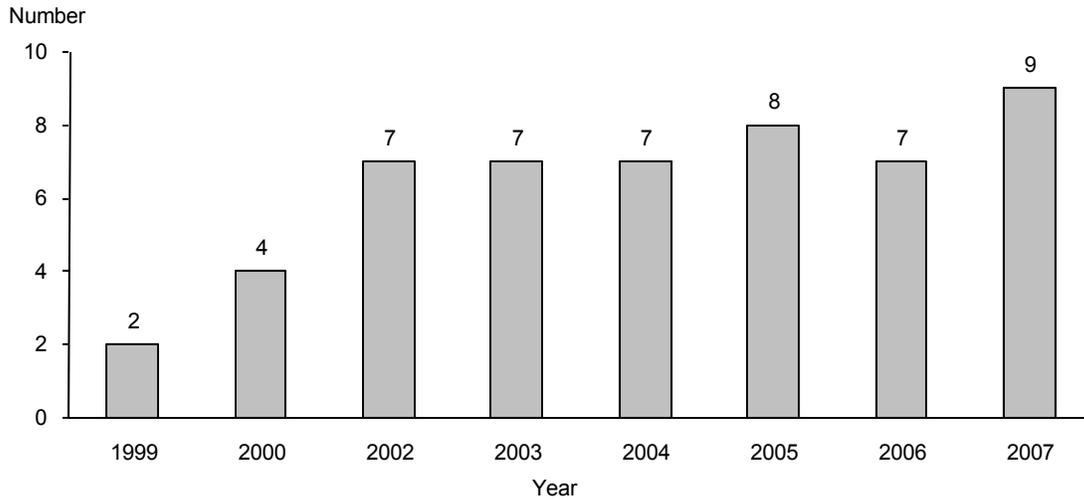
The preliminary data related to postfellowship employment suggest that the target of 40 graduates engaged in research is likely to be successfully met. By the end of summer 2008, the preliminary data noted above suggest approximately 28 of the employed predoctoral fellows are currently in research positions of some type; and approximately 10 postdoctoral fellow graduates are engaged in research. Therefore, in mid-2008, it appears that approximately 38 fellows of the targeted 40 are already engaged in research. However, the preliminary data do not indicate whether or not the research is specific to the field of education. Given the interdisciplinary nature of these training programs, and the various disciplines and departments served by these graduates, it is likely that some of these graduates are not directly engaged in education research. Future data collected need to specifically address the extent to which the postfellowship employment results in active engagement in *education* research by posing questions specific to the nature of their research and future research agendas.

NCES Database Trainings

To what extent do NCES trainings increase the capacity of education researchers to conduct rigorous education research and evaluation?

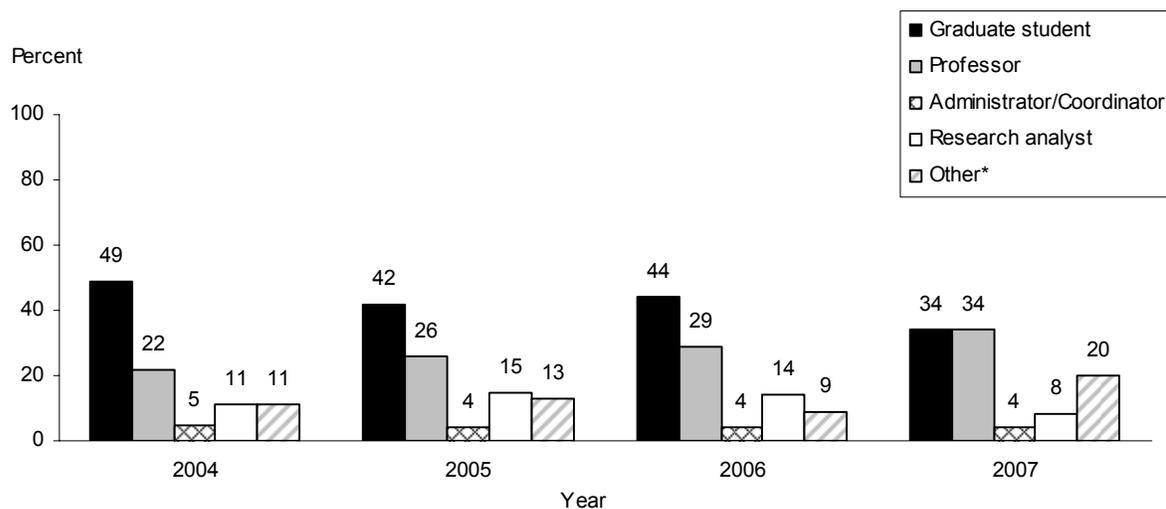
NCES conducted a total of 52 trainings on its various databases between 1999 and 2007. Figure 9 notes the number of different databases for which training was provided for each of these respective years. Data were not available for 2001. However, as indicated in the figure, the data indicate that trainings were offered for substantially more NCES databases after the creation of IES than under OERI. For example, in 1999 and 2000 there were trainings offered for only two and four databases respectively, whereas after the creation of IES the numbers of trainings consistently ranged between seven and nine per year. Unfortunately data were not available regarding numbers of participants served via these trainings. However, completed exit surveys suggest that a minimum of 1,323 persons attended these trainings between 1999 and 2007. However, depending on the response rates for these surveys, it is difficult to get an accurate idea of participation numbers.

Figure 9. Number of NCES database trainings: 1999-2000 and 2002-2007



SOURCE: U.S. Department of Education, National Center for Education Statistics.

The only accessible information on stakeholder participation was the NCES exit surveys from the database trainings. Although it is possible that survey nonrespondents differed in composition from survey respondents, these surveys likely provide relatively reliable data related to overall mix of stakeholders attending NCES database trainings. Figure 10 provides data related to stakeholder participation for 2004 through 2007. As noted in the figure, data indicate a slight decrease over time in the number of graduate students (i.e., 42-49% between 2004 and 2006 versus 35% for 2007), and slight increase in the number of faculty members attending trainings (i.e., steadily increasing numbers starting at 22% in 2004 to 34% in 2007).

Figure 10. Distribution of NCES database trainees by occupation: 2004-2007

SOURCE: U.S. Department of Education, Institute of Education Sciences.

In terms of content, analyses of the 1,323 completed questionnaires from 52 trainings NCES conducted between 1999 and 2007 indicate that the database trainings were deemed by those who attended them to be of extremely high quality across trainings and across years. For example, on average 98 percent of trainees rated the overall quality of training as “good” or “excellent” across all surveyed years. Moreover, in 7 years, a trainee rated seminar overall quality as “poor” in only two instances (of a possible 1,318).

In terms of potential impact on the capacity of the field to use NCES databases to conduct rigorous research, the only related data available are training participants’ responses on the exit surveys regarding their plans to use the NCES database in the future. For 1999 through 2003, participants were asked if they had plans to use NCES datasets within the next year. From 2004 onward, trainees were asked if they planned to use NCES datasets within the next 2 years. As might be expected, nearly all participants had concrete plans for using NCES datasets. Across all years a minimum of 90 percent of participants stated that they planned to use NCES datasets in the future. Approximately one-half of these participants between 2004 and 2007 had previously used a least one NCES database. However the vast majority of these same participants (between 77% and 86%) had not previously published journal articles, doctoral research, books or reports using NCES databases. Unfortunately, no data are available regarding actual usage versus intended usage or plans for using NCES databases.

Other IES Trainings

During the past 2 years, IES has instituted and/or funded other trainings and information sessions aimed at increasing the capacity of the field to conduct rigorous education evaluation. The most intensive of

Rigor

these trainings is the 2-week IES Research Training Institute workshop/training and technical assistance session on cluster randomized trials. The purpose of these summer research training institutes is “to increase the national capacity of researchers to develop and conduct rigorous evaluations of the impact of education interventions.”²² Trainings were provided by Northwestern University, with a grant from IES, during the summers of 2007 and 2008. The Workshop on Evaluating State and District Level Interventions was a 1-day workshop to help states and districts plan and design rigorous evaluations of their policies and programs by providing an overview of quasi-experimental and experimental evaluation designs, with a focus on state-level and district-level design issues. The IES Research Training Institute: Single-Case Design was sponsored by NCSER to increase the capacity of researchers to conduct rigorous special education research using single-case methodologies that incorporate quantitative analyses.

Table 6 provides an overview of these various trainings offered during 2007 and 2008, as well as the number of applicants and numbers of participants. Note that the application and admission process varied across the trainings: the IES Summer Research Training Institute was limited to 30 participants per year and Institute organizers selected participants based on qualifications and likelihood of using the design to conduct rigorous research; state and district level evaluation personnel were invited to the Workshop on Evaluating State and District Level Interventions, but enrollment was unlimited and anyone wanting to participate was accepted; and the IES Research Training Institute: Single-Case Design capped enrollment at 40, but admitted applicants on a first-come, first-served basis. As noted in the table, the demand for the 2-week summer research training institute on cluster randomized trials exceeded capacity: in 2007 there were almost six times more applicants than participant openings; and in 2008 the demand was slightly more than twice the capacity for the training. The extent to which the demand exceeded capacity for this training on cluster randomized trials suggests that there is significant interest in this methodology within the field. Although demand was not as high during the second year of the program, anecdotal evidence suggests that the numbers of applicants during year 2 was substantially lower due to perceptions related to the difficulty of being admitted to the training program.

²² http://ies.ed.gov/ncet/whatsnew/conferences/08rct_traininginstitute/index.asp (retrieved August 7, 2008).

Table 6. Numbers of applicants and participants for Institute for Education Sciences (IES) trainings: 2007–08

Title	Year	Number of applicants	Number of participants
IES Summer Research Training Institute: Cluster randomized trials	2007	178	30
	2008	66	30
Workshop on evaluating state and district level interventions	2008	121	121
IES Research Training Institute: Single-case design	2008	96 ¹	39

¹ After 96 applications were received it was posted on the internet that applications were no longer being accepted. Workshop was first come, first served.

SOURCE: U.S. Department of Education, Institute of Education Sciences.

In addition to these more formal content-based trainings specifically focused on increasing the use of rigorous methodology, IES also offered numerous webinars during 2008 to help increase understanding of the IES grant application, as well as various programs within IES. In total, 12 webinars each lasting approximately 1 to 2 hours were implemented. Topics and dates of these webinars include the following:

Basic Overview Session: IES, NCSER and NCER research topics, the IES goal structure, and peer-review process, 5/7/08, 5/12/08, 5/13/08

Grant Writing Workshop, 5/15/08, 5/20/08

Application Process Session, 5/28/08, 6/30/08

Overview of IES Education Research Training Grants, 7/31/08

Overview of the Evaluation of State and Local Education Programs and Policies Program, 8/1/08, 8/7/08

Grant Writing Workshop for Development Projects (Goal 2), 8/4/08

Grant Writing Workshop for Young Investigators, 8/6/08

Although these webinars were not specifically focused on increasing capacity to conduct rigorous research, these trainings do more indirectly improve the likelihood of researchers being able to conduct rigorous research by increasing understanding of the application and grant writing process, and providing information on specific program areas.

Summary

Quantity and Quality of Rigorous Education Research

To what extent do the research and evaluation studies currently funded by IES meet the highest quality standards related to rigor?

NCER: GPRA data indicate that just prior to the establishment of IES in 2001, 32 percent of funded projects addressing causal questions used randomized experimental designs. Immediately after the establishment of IES there was a drastic increase in the percentage of education research and evaluation projects addressing causal questions that used RCTs, with 82 percent to 100 percent of NCER new research and evaluation projects addressing causal questions using randomized experimental designs.

NCER: In 2007 a new GPRA measure focusing on the evidence standards of the WWC was developed to replace this prior measure. For FY 07, data indicate that 100 percent of new studies of efficacy and effectiveness funded NCER employ research designs that meet evidence standards of the WWC (target was 90%).

NCER: GPRA data related to the percentage of NCER funded research projects that are deemed to be of high quality are questionable in terms of reliability and validity. The discrepancies between the two different methods used to calculate the percentages (FY 01 to FY 04 and FY 05 to FY 07) raise concerns about reliability of ratings over time, and basing the current indicator on the actual review panel scoring process itself limits the meaningfulness of the data since those proposals receiving low scores are not generally funded.

NCEE: To date has 24 large IES-supported evaluation studies currently underway. The number of such evaluations using rigorous methodology in 2000 under the support of OERI was one evaluation study, indicating a significant shift in focus to increasing the rigor of evaluations.

IES's PART program performance data indicate that IES is currently meeting its targeted goals in interventions demonstrating positive effects in reading and writing and enhancing teacher characteristics (6 in 2007, and 3 in 2006 respectively), and exceeding targeted numbers of interventions in mathematics and science interventions (target of 3 in 2007, actual of 4). In addition, the number of interventions increased between 2006 and 2007 for each of the three content areas.

The six interviewed stakeholders representing major education-related organizations strongly believe that IES has increased the quality of research being conducted within the field of education, and that the emphasis on rigor is significantly more pronounced within IES than it was during the era of OERI. Several stakeholders still noted the negative impacts of the strong focus on RCTs, but also stated that the position of IES related to rigorous research and RCTs has been modified over time, and is now more inclusive of other methodologies.

More peer-reviewed publications were published from research grants funded during the first 2 years of IES than from research grants funded during the last 2 years of OERI (93 versus 45). For the 2 years of OERI data (2001 and 2002) versus the 4 years of IES data (2002 through 2005) this translates to an average of 11.3 peer-reviewed publications per year for OERI grants, and an average of 44.5 peer-reviewed publications per year for IES grants.

The Potential to Produce Valid and Rigorous Evidence of Effectiveness

To what extent are these research and evaluation studies producing (or likely to produce) valid evidence?

Analyses of funded proposals for FYs 2004 through 2007 on ten dimensions of high quality research designs indicate that the studies being funded by IES have a high potential for generating rigorous and valid evidence of effectiveness, if the study parameters proposed can be maintained during the study itself, or modified in rigorous ways if necessary.

Analyses indicate that correct power analysis techniques has increased from 71 percent to 81 percent over the 4 years, while the percentage of funded research studies addressing causal questions having an adequate sample size and using the appropriate level and method of randomization has increased from 86 percent to 88 percent, and 86 percent to 94 percent, respectively—indicating that a high percentage of funded studies have these characteristics.

Analyses indicate that the use of valid outcomes has been maintained at a high level over the four years, while the percentage of funded research studies addressing causal questions using systematic data collection procedures for both treatment and control groups has increased 71 percent to 94 percent, and the percentage of studies using longitudinal methods has increased from 57 percent to 88 percent, respectively—indicating that a high percentage of funded studies have these characteristics in 2007.

Analyses indicate that using appropriate analytic methods and properly considering issues of baseline group equivalence has been maintained at a high level over the four years, while the percentage of funded research studies addressing causal questions that appropriately address attrition issues and other threats to study validity has increased from 57 percent to 88 percent, and 57 percent to 81 percent, respectively—indicating that a high percentage of funded studies have these characteristics in 2007.

Capacity of the Field to Conduct Rigorous Education Research

To what extent has IES increased the number and quality of pre- and postdoctoral scientists? To what extent are pre- and postdoctoral scientists funded through IES programs likely to contribute to the quantity and quality of rigorous evidence related to education practice?

Predoctoral and postdoctoral fellowships are the primary mechanisms used by IES for increasing the capacity of the field to conduct rigorous education research. NCER supports fifteen interdisciplinary

Rigor

predoctoral research training programs, including five made in July 2008; and 17 interdisciplinary postdoctoral research training programs. In addition, NCSEER awarded five new grants for Postdoctoral Special Education Research Training Fellowships in July 2008.

NCER has funded a total of 242 predoctoral fellows (2004 through 2008) and 30 postdoctoral fellows (2005 through 2008). Approximately 16.5 percent of the predoctoral fellows (N=40) from 2004 through 2008 are racial-ethnic minorities, slightly lower than national survey statistics from 2005 indicating that approximately 21.4 percent of all doctoral recipients with degrees related education research were racial-ethnic minorities (Hoffer et al., 2006). Approximately 30 percent of the postdoctoral fellows (N=9) from 2005 through 2008 are racial-ethnic minorities, slightly higher than the national survey statistics noted previously.

To date the number of completed Ph.D.s from the predoctoral programs who are employed (including summer 2008 with jobs lined up) is 37, and the number of postdoctoral fellows that have completed the postdoctoral fellowship and were employed by Summer 2008 is ten.

In terms of the PART benchmark data related to the targeted numbers of pre- and postdoctoral scientists IES hopes to impact through its fellowship training programs, although annual output targets were not met for 2007 and 2008, the actual numbers of individuals participating in the IES-funded research training programs were close to the targeted numbers (92% and 94% of the respective targets of 175 and 230). Students participating in the IES predoctoral training programs appear to be of high quality. The average verbal GRE score was 618 (85th to 89th percentile); and the average quantitative GRE score was 695 (68th to 72nd percentile). The predoctoral fellows have substantively higher GRE scores for both the verbal (i.e., 40 percentile points higher) and quantitative sections (37 percentile points higher) than intended education majors, as well as social science applicants and overall graduate school applicants; and also have higher GRE scores (13 percentile points higher on verbal, 19 percentile points higher on quantitative) than successful applicants to the top 20 education doctoral programs in the United States, as identified by Walker (2008).

In terms of the quality of postdoctoral students participating in the fellowships, no extant data related to quality were available for the purposes of this evaluation.

In terms of likelihood of contributing to quantity and quality of rigorous evidence, the following statistics provide some indication of the both the background and experiences of the predoctoral and postdoctoral fellows, as well as their potential to be academically productive: (a) During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES presenting a total of 662 refereed conference presentations, and postdoctoral fellows self-reported 132 refereed conference presentations, and (b) During the 2-year period between 2006 and 2008, predoctoral fellows self-reported to IES having a total

of 126 published/in press papers (excluding conference proceedings), and postdoctoral fellows self-reported 52 published/in press papers.

Given the length of time needed to complete the predoctoral programs, only limited data are currently available related to postemployment. However, an analysis of the data available indicate that approximately 80 percent (N=28) of the employed predoctoral fellows are currently in research positions of some type (i.e., research faculty, postdoctoral fellowships, research position at universities, research position at private research firms, and other noneducation research related). However, it is not possible to determine from the current data what specific fields of research these individuals will pursue.

In terms of the ten postdoctoral fellows that have completed their fellowship and were employed by summer 2008, 50 percent (N=5) obtained tenure-track faculty positions at institutions of higher education, 40 percent (N=4) obtained research positions at universities or university research centers, and 10 percent (N=1) obtained a research position at a private research firm.

The preliminary data related to postfellowship employment suggest that the PART target of 40 graduates engaged in research by 2009 is likely to be successfully met. In mid-2008 it appears that approximately 38 fellows of the targeted 40 are already engaged in research. However, the preliminary data do not indicate whether or not the research is specific to the field of education.

During the past 2 years, IES has also instituted and/or funded the following trainings and information sessions aimed at increasing the capacity of the field to conduct rigorous education evaluation: 2-week Summer Research Training Institute on cluster randomized trials attended by a total of 60 participants(2007 and 2008), 1-day workshop on Evaluating State and District Level Interventions (2008) attended by 121 participants, and a 2-day IES Research Training Institute on single-case design was sponsored by NCSER (2008) attended by 39 participants.

The demand for the Summer Research Training Institute on cluster randomized trials exceeded capacity: in 2007 there were almost 6 times as many applicants as participant openings, and in 2008 demand was slightly more than twice the capacity for the training. The extent to which demand exceeded capacity for this training on cluster randomized trials suggests that there is significant interest in this methodology within the field.

In addition to more formal content-based trainings specifically focused on increasing the use of rigorous methodology, IES also offered numerous webinars during 2008 to help increase understanding of the IES grant application process, as well as various programs within IES. In total, 12 webinars each lasting approximately 1 to 2 hours were implemented.

To what extent do NCSER trainings increase the capacity of education researchers to conduct rigorous education research and evaluation?

Rigor

NCES conducted a total of 52 trainings on its various databases between 1999 and 2007. The data indicate that trainings were offered for substantially more NCES databases after the creation of IES than during OERI. For example, in 1999 and 2000 there were trainings offered for only two and four databases respectively, whereas after the creation of IES the numbers of trainings consistently ranged between seven and nine per year.

Data indicate a slight decrease over time in the number of graduate students (i.e., 42-49% between 2004 and 2006 versus 35% for 2007), and slight increase in the number of faculty members attending NCES trainings (i.e., steadily increasing numbers starting at 22% in 2004 to 34% in 2007).

On average 98 percent of trainees rated the overall quality of training as “good” or “excellent” across all surveyed years. Moreover, in 7 years, a trainee rated seminar overall quality as “poor” in only two instances (of a possible 1,318).

In terms of potential impact on the capacity of the field to use NCES databases to conduct rigorous research, across all years a minimum of 90 percent of NCES training participants stated that they planned to use NCES datasets in the future. Approximately one-half of these participants between 2004 and 2007 had previously used a least one NCES database. However the vast majority of these same participants (between 77% and 86%) had not previously published journal articles, doctoral research, books or reports using NCES databases. Unfortunately, no data are available regarding actual usage versus intended usage or plans for using NCES databases.

RELEVANCE

Based on the established goals and priorities of IES, the evaluation also focused specifically on the impact of the Institute on the relevance, usefulness and timeliness of education research. More specifically, the evaluation addressed the following primary question: To what extent, and in which ways, has IES increased the relevance and timeliness of education research? Given the available timeframe and the scope of work that focused almost exclusively on the use of extant data, the following questions were addressed:

Is IES providing relevant, useful, and accessible data, research and publications to various stakeholder groups? To what extent does this relevance differ among stakeholder groups?

To what extent is IES producing findings (or likely to produce findings) that answer questions that are important; and have practical significance from a policy and practice perspective? To what extent is IES funding research that builds on prior evidence, and using the past to focus on studies most likely to have an impact?

To what extent is IES providing relevant data, and/or funding research and evaluation, that produce relevant findings as defined by the Institute's established priorities?

To what extent is IES producing findings and data in a timely manner that ensures their relevance to current and/or pressing education issues?

The remainder of this section provides available data and findings related to each of these questions. First, the initial three questions related to relevance are addressed; and next, the question related to timeliness is addressed. Finally a brief summary of findings is provided.

Relevance

Relevance of Information

Is IES providing relevant, useful, and accessible data, research and publications to various stakeholder groups? To what extent does this relevance differ among stakeholder groups?

Given the accessible extant data, this question related to relevance, usefulness and accessibility is primarily addressed for NCER and NCSER, as well as NCES. The data related to these questions include Government Performance and Results Act of 1993 (GPRA) data related to the relevance of research projects funded by NCER and NCSER, and NCES data related to relevance from customer satisfaction surveys. Given that groups differ in the types of information they need, and therefore that the inherent

Relevance

relevance and necessity of types of information differ among groups, the heterogeneity of stakeholders should also be addressed to the extent possible when examining relevance. The NCES survey data allow for some disaggregation by stakeholder group to examine the extent to which there were any differences among the various subpopulations. In addition, key stakeholder interviews provided some additional information related to perceptions of the relevance and usefulness of IES data, research and publications. Findings related to each of these data sources are discussed in the remainder of this section.

NCER and NCSEER. Past measures included as part of GPRA have included an indicator related to the relevance of research projects funded by IES. More specifically, this GPRA measure related to relevance was stated as the following:

The percentage of new research projects funded by the Department's National Center for Education Research and National Center for Education Evaluation and Regional Assistance that are deemed to be of high relevance to education practices as determined by an independent review panel of qualified practitioners.

The methodology used for determining the relevance of funded projects consisted each year of selecting a stratified random sample of newly funded research proposals. Table 7 on the next page provides data related to the percentage of newly funded proposals included in the sample for each fiscal year, ranging from approximately one-third of proposals in FY 01 and FY 02 to 87 percent in FY 06. Single page abstracts prepared for each project in the sample were then submitted to panels of practitioners who served as external reviewers. Depending on the year and numbers of proposals, two or three panels were created with each panel reviewing the abstracts for relevance. More specifically, reviewers were asked to “Please rate the overall significance of the proposed research to education in our country. Consider the degree to which the proposed research addresses problems that are of national significance to education and have the potential to contribute to solving those problems.” Each abstract was rated using a 9-point Likert-type scale where 1 represented “very low relevance”, 3 represented “low relevance”, 5 represented “adequate relevance”, 7 represented “high relevance” and 9 represented “very high relevance.” High relevance was defined as having a mean rating of 6.5 or higher on this 1–9 scale.

Table 7. Established target and actual percentages of new National Center for Education Research (NCER) and National Center for Education Evaluation and Regional Assistance (NCEE) projects deemed to be of high relevance: 2001–06

Year	Target	Actual	Approximate Percentage of Total Funded Proposals Included in the Sample ²³
2001	—	21	33%
2002	25	25	33%
2003	37	60	50%
2004	50	50	77%
2005	65	33	65%
2006	75	74	87%

— Not available.

SOURCE: U.S. Department of Education, FY 07 Program Performance Report.

For the purposes of this evaluation, data related to the specific composition of the independent review panel of qualified practitioners were made available for 2001 through 2007. For FY 01 through FY 07 the independent review panels of qualified practitioners generally consisted of a combination of district superintendents and assistant superintendents, school principals, and program managers and directors from state departments of education. For FY 04 and FY 05 the review panels also included senior staff for special education departments within school districts or state departments of education. However, starting in FY 06 these types of staff persons were no longer included in the review panels given that separate reviews for the relevance of NCSER were conducted.

Table 7 indicates the established target and actual percentages of new projects deemed to be of high relevance for each year beginning in FY 01. As noted in the table, there is a sharp increase in perceived relevance between the FY 01 and FY 02 (21% and 25% rated high relevance respectively) and the percentage of projects deemed to be of high relevance starting in FY 03 (between 50% and 74% high relevance, with the exception of 2005). However, it is difficult to know the extent to which the changes in percentages (e.g., 50% to 33% to 74%) represent real changes in the relevance of projects given apparent changes in review panels across the years. Although FY 01 and FY 02 were reviewed by the same panel and at the same time, providing consistency across these two fiscal years, the panels vary across the remaining fiscal years. The types of persons on the independent review panels across these years are similar

²³ Percentages were reported in *Procedure Summary: IES Quality, relevance and research Design GPRA Indicators* provided by IES for all years except 2004. For 2004 the percentage was calculated based on an estimate of 39 funded projects generated by the IES web-based search tool (excluding small business innovation research and training programs as appears to have been done for the other calculations).

Relevance

in terms of position type (e.g., superintendents, principals). However, given that these panel members represent individual opinions as opposed to representatives of national organizations that might be more attuned to education needs and trends on a macro level, even small changes in panel composition are likely to result in vastly different ratings. Nine of the 12 members of the FY 04 review panels were also participants in the 16-member review panels for FY 05, indicating that slightly more than half of participants were consistent across the 2 years. For FY 05 to FY 06 eight of the FY 05 review panel members were also participants in the FY 06 panels consisting of 22 members, indicating only 36 percent of the members participated on the previous review panels. In addition, between FY 05 and FY 06 stakeholders related to special education no longer participated in the same review panels. Given these issues with the reliability and validity of the relevance data, it is difficult to use these data to make inferences about the relevance of research funded by National Center for Education Research and National Center for Education Evaluation and Regional Assistance.

This GPRA measure was discontinued in FY 07 because it included data from evaluation projects that were funded from program appropriations other than the appropriation for Research, Development, and Dissemination (RDD). It was replaced with the following similar measure that includes only data from evaluation projects funded under the appropriation for RDD:

The percentage of new research projects funded by the Department's National Center for Education Research that are deemed to be of high relevance to education practices as determined by an independent review panel of qualified practitioners.

Baseline data for FY 07 were not accessible for the purposes of this evaluation. Similar to the new indicator for NCER, a GPRA indicator specific to research in special education was established in 2006.

This indicator is:

The percentage of new research projects funded by the Department's National Center for Special Education Research that are deemed to be of high relevance by an independent review panel of qualified practitioners.

Baseline data gathered in 2006 indicated 50 percent of the funded NCSER research is highly relevant. The target established for FY 07 is 55 percent. As opposed to the review panels for NCER, panel composition data for NCSER indicate the participation of both individuals within school districts and state departments of education (e.g., district-level resource specialists, directors of special education) as well as representatives from larger national organizations representing special education interests.

NCES. In 1997, 1999, 2001, and 2004 NCES administered a customer survey to help identify areas for improvement in data collection and reporting systems. NCES administered the survey to a random sample of over 3,900 federal policymakers, state policymakers, local policymakers, academic researchers, education association researchers, education journalists and known NCES users. Only those respondents who indicated that they used NCES products were asked about NCES publications and services.

The same four core respondent groups - federal, state and local policymakers and academic researchers – were included in the surveys for 1997, 1999, 2001, and 2004. Therefore, responses for these target groups can be used to examine changes or trends across time. Across these years, survey respondents were asked to rate their level of satisfaction with the relevance of NCES products and NCES services. Satisfaction with the relevance of NCES publications has remained high across the years from 1997 through to 2004, with approximately 90 percent satisfied or very satisfied with the relevance of NCES publications since 1997. Similarly, satisfaction with the relevance of NCES services has remained high, averaging 91 percent across the years 1997 through 2004.

The 2004 survey included additional education stakeholders, allowing for the examination of differences in perceptions across the following subpopulations: policymakers, supervisors/administrators/managers, teachers, researchers or evaluators, reporters/media, and other. Table 8 on the next page provides the data from the 2004 survey by stakeholder group. In general, the different types of users were largely in agreement in their satisfaction with the relevance of information provided in NCES publications, with the exception of reporters/media. Slightly more than three-quarters (76%) of reporters/media were satisfied or very satisfied with the relevance of information as compared to 91 to 98 percent for all other groups. Teachers expressed the highest level of satisfaction, with 98 percent indicating they were satisfied or very satisfied with the relevance of NCES publications. Reporters and members of other media were also least satisfied with the accessibility of information in NCES publications (i.e., “ease of understanding), but the difference between reporters and other stakeholder groups was much smaller.

Relevance

Table 8. Percentage of National Center for Education Statistics (NCES) customer survey respondents satisfied or very satisfied with various aspects of NCES publications and services: 2004

Aspects of NCES publications and services	Supervision, administration, or management						
	All types	Policy-making	Teaching	Research, evaluation or testing	Reporting/media	Other	
NCES publications: Relevance of information	93	93	91	98	93	76	92
NCES publications: Ease of understanding	88	89	87	90	89	83	90
NCES services: Extent to which the information met your needs	94	87	93	99	93	92	95
NCES services: Ease of obtaining the information	88	73	91	89	87	87	89

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 Customer Satisfaction Survey.

In terms of NCES services, as indicated in the table, the different stakeholder groups also indicated general satisfaction with the relevance of NCES services, or more specifically, the “extent to which the information meet your needs.” Across all populations, between 87 percent and 99 percent of respondents stated they were either satisfied or very satisfied with the extent to which information provided by NCES met their needs. Once again teachers had the highest level of satisfaction with 99 percent of teachers stating they were satisfied or very satisfied. Policymakers were least satisfied with 87 percent indicating some level of satisfaction with the relevance of information provided when using NCES services. Policymakers were also least satisfied with the ease of obtaining the information, with 73 percent indicating they were satisfied or very satisfied as compared to an average of 88.5 percent for all other stakeholder groups.

In 2006, NCES changed the methodology for the customer survey. The survey now focuses on collecting data from a random sample of visitors to the NCES website using pop-up windows. Given the change in methodology, data are not comparable to data collected prior to 2006, and these most recent data cannot be reported as part of a trend from the earlier years. Table 9 below shows the percentage of respondents who were satisfied with the relevance of NCES data files and NCES publications. Despite the change in methodology, overall satisfaction rates appear similar to prior survey years: the vast majority of respondents are either satisfied or very satisfied with both the relevance of NCES data files and NCES publications.

Table 9. Percentage of National Center for Education Statistics (NCES) customer survey respondents satisfied or very satisfied with relevance: 2006–07

Year	NCES data files	NCES publications
2006	94	95
2007	94	94

SOURCE: U.S. Department of Education, National Center for Education Statistics.

Key Stakeholder Perceptions. As noted previously, SEI/CEEP conducted key stakeholder interviews to supplement extant data. Interviews were completed with key stakeholders from the following organizations and associations: American Educational Research Association (AERA), American Psychological Association (APA), National Academy of Sciences, Council of the Great City Schools, Knowledge Alliance, and National Sorority of Phi Delta Kappa. Interpretation is somewhat limited due to the small numbers of education-related organizations represented. However, the stakeholders included in the data represent some of the largest and most representative education-related organizations in the nation (e.g., AERA and APA); and the interview responses represent these persons' perceptions of the views and opinions of their broader constituencies, rather than the individual opinions of six persons. Therefore, the data from these six interviews do provide some valuable insight into perceptions of IES impact, particularly when interpreted within the context of other available data.

The interviewed stakeholders generally believed that IES should get “good marks” in relevance, with most also sharing the opinion that relevance has only recently become a focus of IES. One stakeholder noted that relevance is the “new key word” at IES, and previously the emphasis appeared to be almost exclusively on rigor. Although most of these stakeholders discussed the recent emphasis on relevance with a relatively neutral stance, one person did give IES “a C or C+” on relevance due to this perceived late attention to this criterion. This person criticized IES for both the length of time it took to begin thinking about relevance, as well as the perception that “rigor trumps relevance at IES,” whereas this stakeholder believed both should be considered equally in funding decisions. This person believes that the emphasis on rigor over relevance has resulted in some of the relevant topics not receiving the attention that they need.

Some stakeholders also specifically noted a perceived difference in the relevance of research being funded by OERI versus IES, stating that IES appears to be better than OERI in its ability to tie research to the field. As expressed by one stakeholder representing an educational association:

“Russ has done better than his predecessors in trying to connect the research field with the practitioner world in a way that others haven’t, but the connection is still so new and fragile ... IES has a much better understanding of why research needs to be connected.”

Some interviewees noted that the increased collaborative relationships with other education organizations such as the Council of Great City Schools are likely to further increase the relevance of IES research and activities.

Relevance

Significance of Information

To what extent is IES producing findings (or likely to produce findings) that answer questions that are important; and have practical significance from a policy and practice perspective? To what extent is IES funding research that builds on prior evidence, and using the past to focus on studies most likely to have an impact?

A key indicator of relevance is the degree to which research results are made accessible to diverse groups of stakeholders. The WWC has expended considerable effort to classify educational research with respect to rigor, and to summarize those findings in areas of importance for educational practice.

For studies that meet the evidence standards, intervention reports are prepared and disseminated via the WWC website. From August 2006 through July 2008, the WWC produced and released 84 Intervention Reports across the topics of beginning reading, character education, dropout prevention, early childhood education, English language learning, and elementary and middle school mathematics, with a number of interventions under review at the current time. Of these 84 interventions, the WWC determined that 63 demonstrated positive or potentially positive effects in at least one outcome domain related to student achievement. The WWC also produced and released six Topic Reports (elementary and middle school math, English language learning, beginning reading, dropout prevention and character education) with a seventh topic report on early childhood education forthcoming in late summer 2008. The topic reports summarize the findings across all intervention reports within a topic area. These topic reports are clearly and concisely written summaries of the effectiveness evidence in these areas across all of the intervention reports issued. They consist of many graphic presentations of complex research findings and standardized metrics for understanding the impact of interventions in a particular area.

In addition, the Clearinghouse also produced a series of 10 quick review documents in 2008. According to the WWC website, the

“What Works Clearinghouse (WWC) quick reviews are designed to provide education practitioners and policymakers with timely and objective assessments of the quality of the research evidence from recently released research papers and reports whose public release is reported in a major national news source. These reviews focus on studies of the effectiveness of education or school-based interventions serving students in the prekindergarten through twelfth grade age range, as well as those in a postsecondary setting.”

To further assist practitioners in applying rigorously tested research strategies WWC has published a set of four practice guides. These guides have been designed to provide an overview of strategies to address the following educational issues:

Turning Around Chronically Low-Performing Schools

Encouraging Girls in Math and Science

Organizing Instruction and Study to Improve Student Learning

Effective Literacy and English Language Instruction for English Learners in the Elementary Grades

The goal of these different types of publications is to increase the accessibility and relevance of the research for the every-day practitioner as well as to assist researchers in building upon previously validated research in their work. While the face validity of the relevance of these publications is obvious, it would be useful in future evaluations to survey users about the relevance and usability of the publications to make a stronger judgment of relevance.

IES Priorities

To what extent is IES providing relevant data, and/or funding research and evaluation, that produce relevant findings as defined by the Institute's established priorities?

A proposed set of priorities for IES was developed in July 2005 by IES Director Whitehurst; and modified following publication in the *Federal Register* and the solicitation of public comment. The Institute priorities were approved by the National Board for Education Sciences during the September 2005 Board Meeting. The priorities included the following:

By providing an independent, scientific base of evidence and promoting and enabling its use, the Institute aims to further the transformation of education into an evidence-based field, and thereby enable the nation to educate all of its students effectively.

In pursuit of its goals, the Institute will support research, conduct evaluations, and compile statistics in education that conform to rigorous scientific standards, and will disseminate and promote the use of research in ways that are objective, free of bias in their interpretation, and readily accessible.

The four goals associated with these established priorities were the following:

To develop or identify a substantial number of programs, practices, policies, and approaches that enhance academic achievement and that can be widely deployed;

To identify what does not work and what is problematic or inefficient, and thereby encourage innovation and further research;

Relevance

To gain fundamental understanding of the processes that underlie variations in the effectiveness of education programs, practices, policies, and approaches; and

To develop delivery systems for the results of education research that will be routinely used by policymakers, educators, and the general public when making education decisions.

IES's plan for addressing its research priorities focuses on the following three elements: (1) assessing whether the Institute is providing opportunities for researchers to obtain funding for work on each of the topics identified in the priorities, (2) assessing whether the mix of grant applications within each topic is appropriate to the Institute's goals of determining what works and what doesn't, as well as reasons for variations in program effectiveness, and (3) assessing whether the yield of grants within each topic is advancing the Institute's goals, particularly the goal of developing and identifying programs and practices that are effective in enhancing academic achievement. Based on these assessments a determination is made as to whether new research activities should be created, or existing opportunities modified, to best address the Institute's priorities. This evaluation relied on this same general process established by the Institute to examine the extent to which IES is providing relevant data, and/or funding research and evaluation, that produce relevant findings as defined by the Institute's established priorities.

Do Research Opportunities Fit the Priorities?

The NBES 2006 Annual Report notes that "it should not be surprising that most of the conditions and outcomes identified in the priorities are presently being covered" given that the priorities published in the fall of 2005 were built on the research competitions that were already in place. However, the Institute appears to have effectively used its overall framework for its research grant programs (i.e., organizing programs within NCER and NCSER by outcomes such as reading or mathematics; type of education condition such as curriculum and instruction, teacher quality, administration, systems, and policy; grade level; and research goals) and self-assessment process to identify gaps in existing research opportunities. For example, the NBES 2006 Annual Report notes that a new program on early intervention, early childhood and special education and assessment for young children with disabilities was developed to help rectify a notable gap in coverage in special education research on infants and toddlers with disabilities. In addition, the Annual Report notes that a postsecondary education research program was launched to address some areas of postsecondary research identified in the priorities but underrepresented in existing research programs.

In the time since the 2006 Annual Report, IES has continued to create new programs, or modify existing programs, to fill gaps in the priorities that are not covered by existing research programs. For example, these programs for NCER include the following:

Early Childhood Programs and Policies was created for FY 08 to contribute to improvement of school readiness skills (e.g., prereading, early mathematical skills, language, vocabulary, social skills) of prekindergarten children. Although the Institute funded several early childhood projects in early literacy, early mathematical skills, and teacher quality through previously existing programs (e.g., Reading and Writing, Mathematics and Science Education, Teacher Quality), the Institute launched this program to attract more proposals related to early childhood research and policies.

Education Technology was launched for FY 08 to increase the quantity and quality of rigorous research being conducted to develop and evaluate new education technology tools and evaluate existing education technology products. Although the Institute funded some technology projects through its other competitions, this new program was created to call attention to the current gap in research related to education technology.

Social and Behavioral Context for Academic Learning was implemented for FY 08 to support research on interventions designed to improve social skills and behaviors that support academic and other important school-related outcomes (e.g. attendance, high school graduation rates) in typically developing students from kindergarten through Grade 12.

High School Reform was launched in FY 06 to examine the effectiveness of different high school reform practices on student outcomes. The program was designed to support crosscutting reform efforts to complement existing research programs on teacher quality, reading and writing, interventions for struggling adolescent and adult readers, mathematics and science education, education leadership, and policy and systems. However, the creation of this new program helps to ensure that these grade levels are not overlooked in the research portfolio. For similar reasons, for FY 09 the program was modified to also include middle school reform, becoming Middle and High School Reform.

Is the Mix of Research Appropriate?

Although the Institute funds research in seven categories (i.e., identification, development, efficacy, scale-up/effectiveness, assessment and tools/measurement, training, and centers), four of these categories provide “a logical and progressive ordering of research activities towards the goal of developing and identifying programs and practices that are effective in enhancing academic achievement” (NBES, July 2006, pp. 21). Goal 1, identification, focuses on identifying existing programs, practices, and policies that are differentially associated with student outcomes and the factors that mediate or moderate the effects of these

Relevance

programs, practices and policies. Goal 2, development, focuses on developing programs, practices, and policies that are potentially effective for improving outcomes. Goal 3, efficacy, focuses on establishing the efficacy of fully developed programs, practices or policies that either have evidence of potential efficacy or are widely used but have not been rigorously evaluated. Goal 4, scale-up/effectiveness, focuses on providing evidence on the effectiveness of programs, practices and policies implemented at scale.

In examining the mix of research IES has used these four goals to assess the extent to which the mix of grant applications within each topic is appropriate to the Institute's goals of determining what works, what doesn't, and understanding the processes that underlie variations in program effectiveness. According to the 2006 NBES Annual Report (pp. 21),

“While there is no formula for determining the appropriate mix of research across these categories, the Institute wants to see a distribution that has the shape of a triangle, with the base consisting of identification and development activities, the second level representing small-scale field tests, and the apex representing evaluations of programs and practices at scale.”

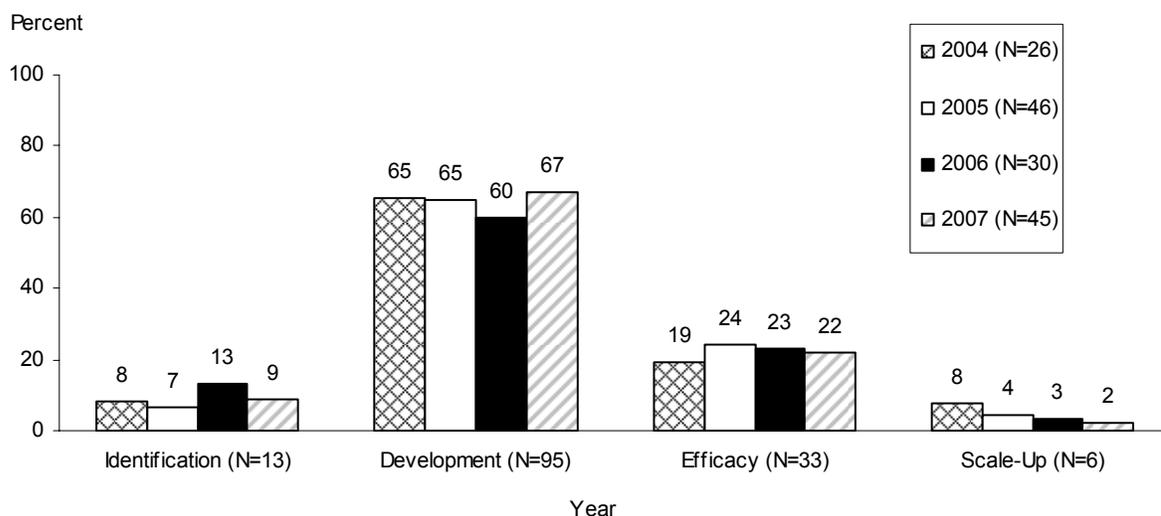
The assessment presented in the 2006 NBES Annual Report concludes that the distribution of overall grant applications received across Goals 1 through 4 since 2002 has the desirable triangular shape, with identification and development representing 60 percent of applications (i.e., 11% identification, 49% development), small field tests at 31 percent, and evaluations of programs implemented at scale at 9 percent. The goal is to have both sufficient upstream work to generate a new generation of programs and practices, while also having sufficient downstream work in moving interventions to scale and evaluating their effectiveness.

For the purposes of this evaluation, the more meaningful question related to the mix of grants is an examination of funded research grants as opposed to grant applications received. While analyzing the mix of grant applications provides some useful information, this assessment that focuses on grant applications addresses the extent to which IES is generating interest in funding opportunities across the four goals. Analyzing the mix of funded research moves beyond interest and opportunities to provide a more meaningful assessment of the extent to which the overall funded research program is balanced in terms of the mix of funded programs.

Across all grant years from 2004 through 2007, the years during which these four goals were used for funding purposes, the following is the distribution of funded NCER grants: 8.8 percent Identification (N=13), 64.6 percent Development (N=95), 22.4 percent Efficacy (N=33) and 4.1 percent Scale-

Up/Effectiveness (N=6). Figure 11 below also illustrates the distribution across FY 04 through FY 07. As noted in the figure, the distribution across these grant years was relatively stable in terms of goals, with the following slight exceptions: a slightly greater percentage of identification NCER grants funded in 2006 (13% versus 7-9%), a slightly lower percentage of development NCER grants funded in 2006 (60% versus 65-67%), and a slightly greater percentage of scale-up studies funded in 2005 (8% versus 2-4%).

Figure 11. Distribution of NCER funded research by goal category and year: 2004-2007



NOTE: Detail may not sum to totals due to rounding.
 SOURCE: U.S. Department of Education, Institute of Education Sciences.

Although not technically classified as efficacy grants given that these goal categories did not exist during FY 02 and FY 03, it should be noted that during these funding years a much greater percentage of efficacy studies were funded. During these years 22 randomized controlled trial design studies were funded: the Preschool Curriculum Evaluation Research Program funded 14 efficacy studies in FY 02 to FY 03, and the Social and Character Development Program funded 8 efficacy studies in FY 03. Although technically not labeled as Goal 3 efficacy studies, the Preschool Curriculum Evaluation Research Program and the Social and Character Development Program are categorized as such for the purposes of these analyses since they are comparable to those studies funded under the efficacy goal. Therefore, these two specific programs represent 40 percent of the total number of all efficacy studies (N=55) funded by NCER between FY 02 and FY 07.

As noted previously, the Institute has stated a desire for a “distribution that has the shape of a triangle” in order to ensure both sufficient upstream work to generate a new generation of programs and practices, and sufficient downstream work in moving interventions to scale and evaluating their effectiveness. As indicated by the above data, the mix of funded NCER research grants for the most recent IES years does

Relevance

still resemble a triangle with more identification and development activities, fewer small-scale field tests, and practices at scale at the apex. However, the base of identification and development grants is wider than the overall mix of grant applications noted in the 2006 NBES Annual Report as having the desirable triangular shape, and also wider than the earlier years of funded research by the Institute. The 2006 Annual Report noted the following overall distribution: 60 percent identification and development, 31 percent efficacy or small field tests, and 9 percent scale-up/effectiveness or programs implemented at scale. In contrast, the average percentage of identification and development grants funded in 2006 through 2007 is 70.9 percent (N=56), representing a much broader base for the triangle as compared to 60 percent; and the average percentage of scale-up grants funded in FY 06 through FY 07 is 2.5 percent (N=2), representing a smaller apex of the triangle as compared to 9 percent.

The evaluation also examined the distribution across the various goals for each of the program content areas from 2002 through 2007.²⁵ Table 10 provides the relevant data. As noted in the table, between 2002 and 2007, the largest percentage of grants has been awarded to the following content areas: cognition and student learning (25.8%, N=56), reading and writing (21.7%, N=47), and mathematics and science education (15.2%, N=33). In terms of the mix of funded research within each of these content areas, the findings include the following:

The majority of content areas, with the exception of High School Reform and Education Policy, Finance and Systems are making virtually no use of the identification goal that focuses on identifying existing programs, practices, and policies that are differentially associated with student outcomes and the factors that mediate or moderate the effects of these programs, practices and policies. There is an absence of scale-up research within the vast majority of content areas. For example, although the two Teacher Quality grant programs (i.e., Mathematics and Science Education and Reading and Writing) have funded a combined total of 37 grants, not a single scale-up grant has been awarded in either program. Between FY 02 and FY 007 a total of six scale-up grants have been awarded across all NCER content areas.

²⁵ For the purposes of this analyses, measurement/assessment and “No Goal” were also included, as well as years FY 02 and FY 03. Therefore, percentages for the four core goals discussed above may not match the figures and tables above. See notes below the table for data sources.

Table 10. National Center for Education Research (NCER) funded research distribution across the various goals for each of the program content areas: 2002–07

Content area	Total number of grants	Content area as percentage of total number of grants	Identification		Development		Efficacy		Scale-Up		Measurement		No Goal	
			Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent	Num-ber	Per-cent
Total	217	†	11	5.1	116	53.5	56	25.8	6	2.8	19	8.8	9	4.1
Cognition and Student Learning	56	25.8	0	0.0	43	76.8	4	7.1	†	0.0	1	1.8	8	14.3
Education Leadership	5	2.3	1	20.0	3	60.0	1	20.0	0	0.0	0	0.0	0	0.0
Education Policy, Finance and Systems	10	4.6	5	50.0	2	20.0	2	20.0	0	0.0	1	10.0	0	0.0
High School Reform	7	3.2	3	42.9	0	0.0	3	42.9	1	14.3	0	0.0	0	0.0
Mathematics and Science Education	33	15.2	0	0.0	20	60.6	7	21.1	3	9.1	2	6.1	1	3.0
Preschool Curriculum Evaluation Research	14	6.5	0	0.0	0	0.0	14	100.0	0	0.0	0	0.0	0	0.0
Reading and Writing	47	21.7	1	2.1	28	59.6	7	14.9	2	4.3	9	19.1	0	0.0
Social and Character Development	8	3.7	0	0.0	0	0.0	8	100.0	0	0.0	0	0.0	0	0.0
Teacher Quality: Mathematics and Science Education	16	7.4	0	0.0	10	62.5	4	25.0	0	0.0	2	12.5	0	0.0
Teacher Quality: Reading and Writing	21	9.7	1	4.8	10	47.6	6	28.6	0	0.0	4	19.0	0	0.0

† Not applicable.

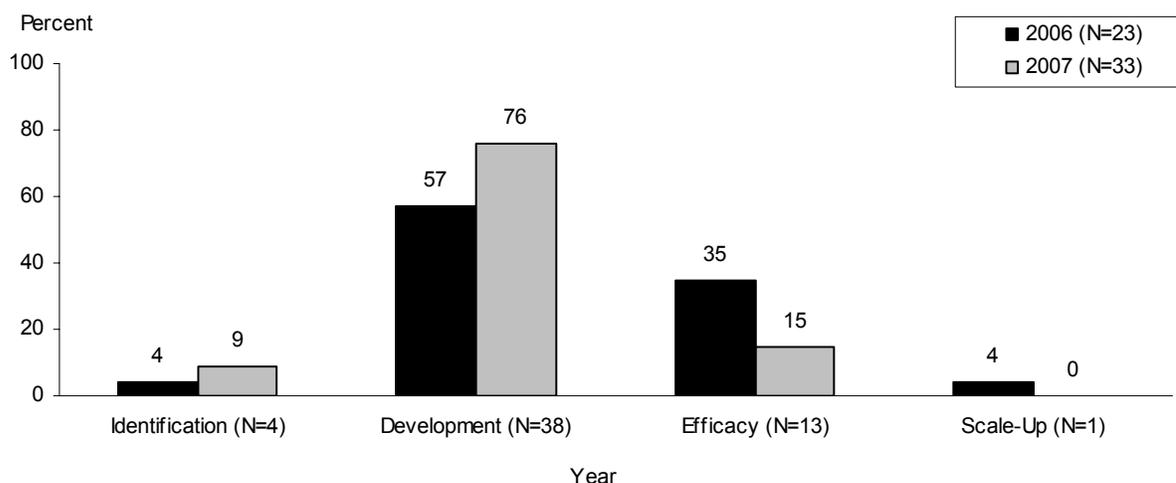
NOTE: Although goal categories were not used for FY 02 and FY 03, the National Center for Education Research: Projects and Programs 2002-2007 publication notes the category that would be most applicable for some grants. For the following programs the "No Goal" was reclassified as follows using the categories noted in the publication: for Cognition and Student Learning 11 "No Goal" grants were classified as development, for Reading and Writing 13 "No Goal" were reclassified as 11 development and 2 measurement, for Teacher Quality: Mathematics and Science 1 "No Goal" was reclassified as development, and for Teacher Quality: Reading and Writing 3 "No Goal" were reclassified as 2 development and 1 efficacy. In addition, the web-based search tool generates the incorrect numbers of Preschool Curriculum Evaluation Research and Social and Character Development grants; given the strong documentation of the correct number of grants in other sources, the numbers in the chart were included based on multiple other data sources.

SOURCE: U.S. Department of Education, Institute of Education Sciences, Web-based Search Tool at <http://ies.ed.gov/funding/grantsearch/index.asp>. Retrieved September 25, 2008.

Relevance

The relatively low numbers of funded NCER efficacy studies are particularly surprising in key, long-standing content areas such as Reading and Writing (7 efficacy studies between 2002 and 2007). Efficacy studies focus on establishing the efficacy of fully developed programs, practices or policies that either have evidence of potential efficacy or are widely used but have not been rigorously evaluated. The low numbers of efficacy studies in this key content area is not likely due to a lack of fully developed reading or mathematics programs in need of field tests of their efficacy. For example, in 2006-2007 the WWC examined 887 studies of 153 beginning reading programs. For beginning reading, 24 intervention programs met evidence standards, suggesting that the remaining 129 beginning reading programs identified by WWC represent a large population of programs needing to be rigorously examined via field-studies of efficacy. An additional six efficacy studies have been funded through the Teacher Quality: Reading and Writing Program. However, these grants have a slightly different focus and do not detract from the relatively low numbers of reading and writing efficacy studies that have been funded from 2002 through 2007.

Similarly, funded research by goal category can be examined to look at the mix of research within NCSER. Figure 12 on the next page provides the details related to the percentage of NCSER projects funded for 2006 and 2007 for each of the primary categories: identification, development, efficacy and scale-up. As noted in the figure, the overall pattern of funding is similar to NCER in terms of the majority of funded projects falling within the development category that focuses on developing programs, practices, and policies that are potentially effective for improving outcomes. Across the 2 years approximately 68 percent of all NCSER research grant funding was allocated to development, as compared to 64.6 percent for NCER research grant funded between 2004 and 2007. For both NCER and NCSER approximately 7 percent of funding was allocated for identification projects. Approximately 25 percent of funded grants fall within the other two aggregated categories for each Center (i.e., efficacy and scale-up).

Figure 12. Distribution of NCSER funded research by goal category: 2006-2007

NOTE: Detail may not sum to totals due to rounding.

SOURCE: U.S. Department of Education, Institute of Education Sciences, Web-based Search Tool at <http://ies.ed.gov/funding/grantsearch/index.asp>. Retrieved September 25, 2008.

Is the Research Yielding Findings That Will Enhance Academic Achievement?

Given that final research reports are not yet available for the majority of IES studies, grant project descriptions from the archived OERI website and the IES website were reviewed to examine the likelihood they would yield findings related to academic achievement. Studies were classified as addressing student achievement outcomes if they explicitly stated that their outcome measures were standardized achievement measures, validated measures of specific learning outcomes, or end-of-course assessments. Studies that had cognitive processing outcomes only (such as memory, metacognition, and problem solving) or student affective or behavioral characteristics (e.g., engagement or attitudes toward learning) only were classified as NOT addressing student achievement.

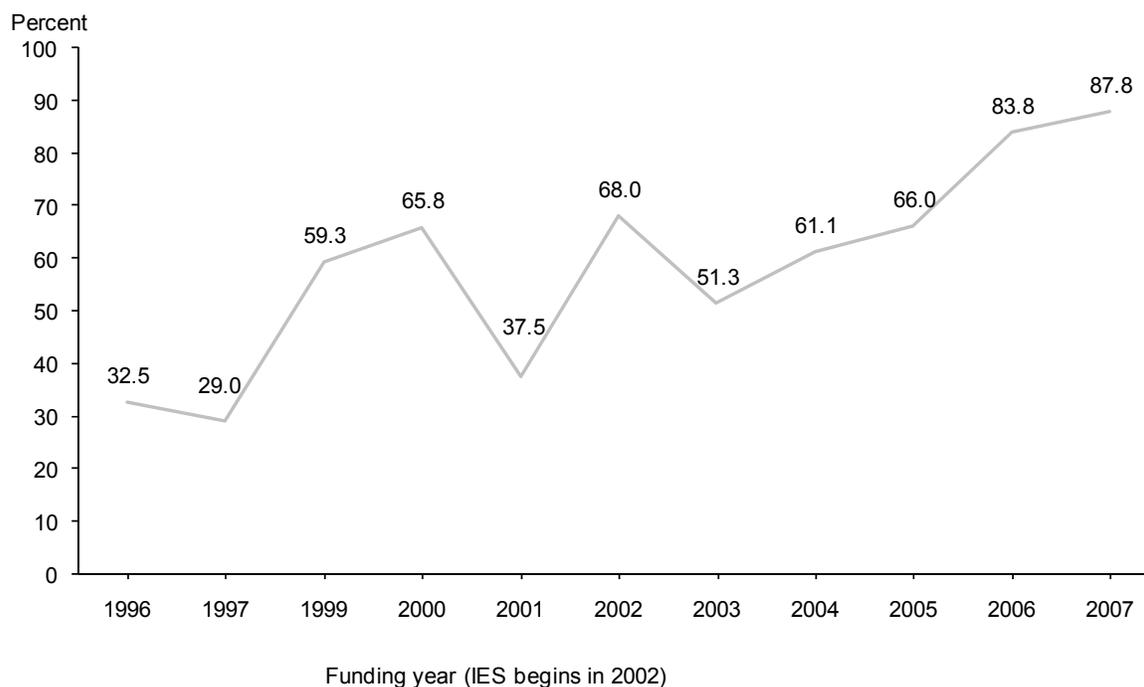
A total of 179 OERI studies and 231 IES studies²⁶ were classified. Figure 13 shows the percentage of studies classified as having achievement outcomes from 1996 to 2007 (IES began in 2002). There are considerably more proposals funded in the IES competitions that address student achievement outcomes

²⁶ Our numbers reflect studies funded under NCER. We do not include Small Business Innovation Research, National Research and Development Centers, Predoctoral and Postdoctoral Training Programs, or Unsolicited and Other Awards grant programs.

Relevance

than under OERI (mean OERI 44.8% versus mean IES 69.7%, representing nearly a 25% increase). In addition there has been a steady increase in the percentage of IES NCER studies that have addressed student achievement outcomes (a 19.8% increase from 2002 to 2007).

Figure 13. Percentage of funded studies with student achievement outcomes: 1996-1997 and 1999-2007



SOURCE: SEI/CEEP analyses of IES data.

Timeliness

To what extent is IES producing findings and data in a timely manner that ensures their relevance to current and/or pressing education issues?

The timeliness of data, research and publications is also critical to relevance and usefulness. Therefore, to the extent possible with the available data, the evaluation also examined the extent to which IES is producing findings and data in a timely manner that ensures their relevance to current and/or pressing education issues. First, NCES data related to both the timeliness of its data and the timeliness of its publications are provided in terms of analyzing release dates; as well as NCES data related to customer perceptions of the timeliness of NCES data, publications and services. Next, timeliness data related to the OMB clearance process for REL randomized controlled trial studies are discussed. In addition, data related

to the timeliness of findings from NCER funded grants are provided. Finally, a brief overview is provided of some strategies and initiatives employed by the Institute to help ensure it is producing findings and data in a timely manner.

NCES

NCES has embedded within its infrastructure numerous measures of timeliness to ensure that the Center produces national databases within reasonable timeframes, and has focused its efforts since 2003 on reducing turnaround of database releases.²⁷ For example, in 2005 NCES established the following timeliness goals:

In 2006, 90 percent of initial releases of data will occur (a) within 18 months of the end of data collection or (b) with an improvement of 2 months over the previous time of initial releases of data from that survey program if the 18-month deadline is not attainable in 2006. In 2007 through 2010, NCES will reduce by 2 months each year the deadline for initial release until the final goal of 12 months is reached (i.e., 16 months in 2007, 14 months in 2008 and 12 months in 2009 and beyond).

As shown in table 11, for both 2006 and 2007, NCES met or exceeded its target in terms of timeliness of data releases. For 2006, the percentage of NCES statistics program initial releases that either met the target number of months (18), or showed at least a 2-month improvement over the prior release, was 90 percent. In 2007, NCES exceeded its target with all 20 initial releases meeting their target release dates: 16 of the 20 reports (80%) were released in 16 months or less, and the remaining four had a reduction of 2 or more months in the time from end of data collection to release when compared to the prior administration of the survey. For these four data releases, the range of reduction was from 7 to 19.5 months.

²⁷ For example, the PART assessment notes that beginning in 2003 NCES Program Improvement Plans included a “focus on improving the timeliness of NCES products and services”; and states that the status is “completed.” The IES Director made improving the timeliness of release of information from NCES surveys a priority and established a performance measure to track time-to-release of survey results.

Relevance

Table 11. Timeliness goals for release of National Center for Education Statistics (NCES) data: 2006–09

Year	Target	Actual
2006	90	90
2007	90	100
2008	90	(¹)
2009	90	(¹)

¹ To Be Determined

SOURCE: U.S. Department of Education, National Center for Education Statistics.

NCES also set specific goals for the timeliness of National Assessment of Educational Progress (NAEP) data for Reading and Mathematics Assessment in support of No Child Left Behind, and regularly tracks this data every 2 years. The NAEP timeliness data are operationalized as the interval between the end of the applicable data collection cycle and submission of the corresponding “first release publication” to the National Assessment Governing Board (NAGB). The goals were established to help ensure that NAEP results are available within 6 months of each reading and mathematics assessment. As shown in table 12 below, the number of months from the end of data collection to the initial release of results was 8 months during the baseline year (2003), greater than the established goal of 6 months. However, in 2005 the targeted goal of 6 months was met, and in 2007 the number of months from the end of data collection to initial release of the results was 5.25 months, approximately 3 weeks less than the targeted goal.

Table 12. Number of months from the end of National Assessment of Educational Progress (NAEP) Reading and Mathematics Assessment data collection to initial release of the results: 2003–09

Year	Target number of months	Actual number of months
2003	6	8.00
2005	6	6.00
2007	6	5.25
2009	6	(¹)

¹ To Be Determined

SOURCE: U.S. Department of Education, National Center for Education Statistics.

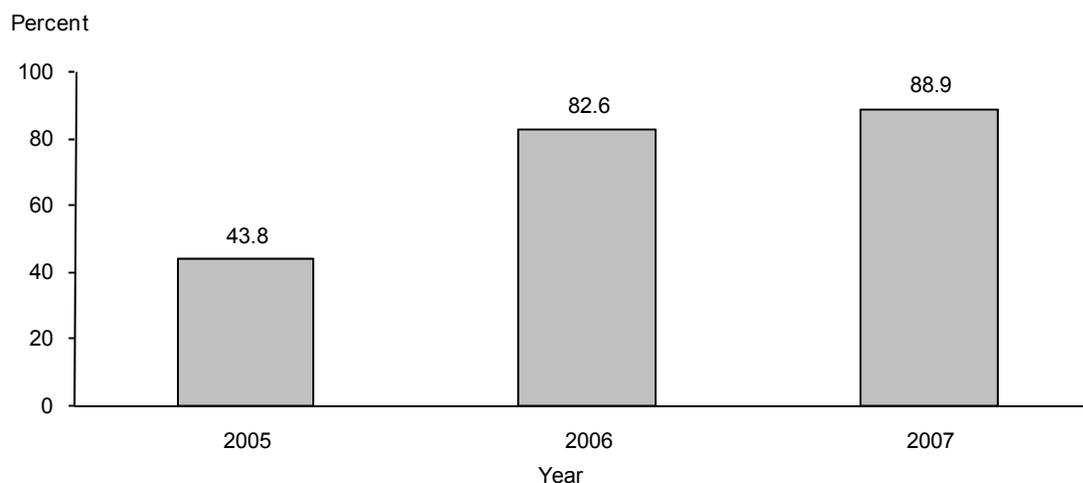
Finally, NCES also provided internal, departmental timeliness data²⁸. The baseline measure for 2005 indicated for the 32 reports that year, the average number of months to release was 19.8 (SD = 8.4), with a minimum of 4 months and a maximum of 39.5 months. Based on its review of the baseline data, NCES set itself the standard for FY 06 of submitting all first release publications within 18 months of the end of data

²⁸ Baseline NCES data and the first year of reporting consisted of combined NCES and NAEP data, whereas subsequent years reported timeliness data separately for NCES and NAEP.

collection, or, if not within 18 months, then at least 2 months faster than baseline production rates. For 2006, 21 out of 23 reports (91%) met the combined goal; and 19 of the 23 (82.6%) were released within 18 months as compared to the baseline in 2005 where 14 of the 32 reports (43.8%) were released within 18 months. For 2006, the average number of months to release was 14.4 (SD = 5.0), with a minimum of 9.2 months and a maximum of 28 months.

For 2007, the NCES target production rate was decreased to 16 months, with two criteria for meeting goals in 2007: first release within 16 months of end of data collection, or if not within 16 months, then there was at least a 2-month reduction over previous release time. For 2007, there were 26 out of 27 reports (96%) that met the combined goal; and 22 out of the 27 reports (81.5%) were released within 16 months. For 2007, the average number of months to release was 12.3 (SD = 5.2), with a minimum of 6.8 months and a maximum of 29.9 months. To provide a comparison across the years using a common measure (i.e., as opposed to the NCES goal that changes to reflect decreasing production times each year), figure 14 below provides the percentage of NCES publications released in 18 months or less from 2005 through 2007. As indicated in the figure, the percentage of NCES publications released within 18 months or less from the end of applicable data collection has increased each year, with a significant change from 2005 to 2006.

Figure 14. Percentage of NCES publications released in 18 months or less: 2005-2007



SOURCE: U.S. Department of Education, National Center for Education Statistics.

In addition to these time-to-release measures, NCES also gathers data related to customers’ perspectives on the timeliness of NCES data, publication, and services. As noted previously, in 1997, 1999, 2001, and 2004 NCES administered a customer survey to a random sample of more than 3,900 federal policymakers, state policymakers, local policymakers, academic researchers, education association researchers, education

Relevance

journalists and known NCES users. Given that the same four core respondent groups—federal, state and local policymakers, and academic researchers—were included in the surveys for 1997, 1999, 2001, and 2004 these responses for these target groups can be used to examine changes or trends across time.

As indicated in table 13, satisfaction with the timeliness of NCES databases has increased over time from 52 percent in 1997 to 78 percent in 2004. Satisfaction with the timeliness of NCES publications has ranged from 72 percent in 1997 to 77 percent in 1999 and 2004. Satisfaction with the timeliness of NCES services remained high across the survey years, averaging 89 percent. Given that the definition of “services” has varied over the years (e.g., the 2001 survey separated out questions related to the NCES website as opposed to prior surveys that included this as part of “services”) may mean that differences between years may be due to the changing definition of services.

Table 13. Percentage of survey respondents that were satisfied or very satisfied with the timeliness of National Center for Education Statistics (NCES) data files, publications and services: 1997–2004

Year	NCES data files		NCES publications		NCES services	
	Target	Actual	Target	Actual	Target	Actual
1997	Baseline	52	Baseline	72	Baseline	89
1999	85	67	85	77	85	93
2001	90	66	90	74	90	88
2004	90	78	90	78	90	84

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 Customer Satisfaction Survey.

The 2004 survey also included additional education stakeholders, allowing for the examination of differences in perceptions across the following subpopulations: policymakers, supervisors/administrators/managers, teachers, researchers or evaluators, reporters/media, and other. As with most of the measures regarding satisfaction, teachers expressed the highest level of satisfaction with the timeliness of NCES publications (96%), and were second highest after supervisors, administrators and managers for timeliness of the release of NCES data (76% teachers, 79% supervisors, administrators and managers). As noted in table 14 below, for both NCES publications and NCES data, policymakers and reporters/members of the media reported the lowest levels of satisfaction.

Table 14. Percentage of National Center for Education Statistics (NCES) customers that were satisfied or very satisfied with the timeliness of NCES publications by stakeholder group

Product/Service	All types	Policy-making	Supervision, administration, or management	Teaching	Research, evaluation or testing	Reporting/media	Other
NCES publications	78	54	79	96	70	63	77
Release of NCES data	70	61	79	76	66	49	67

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 Customer Satisfaction Survey.

As noted previously, in 2006 NCES changed the methodology for the customer survey to a random sample of visitors to the NCES website using pop-up windows. Given the change in methodology, the data are not comparable to data collected prior to 2006, and this most recent data cannot be reported as part of a trend from the earlier years. Table 15 shows the percentage of respondents who were satisfied with the timeliness of NCES data files, NCES publications, and NCES services. Although cause-and-effect cannot be determined, it appears that the current methodology results in reports of higher levels of satisfaction than the prior methodology.

Table 15. Percentage of National Center for Education Statistics (NCES) customers that were satisfied or very satisfied with the timeliness of NCES data files, publications and services: 2006–07

Year	NCES data files		NCES publications		NCES services	
	Target	Actual	Target	Actual	Target	Actual
2006	90	86	90	85	90	92
2007	90	84	90	86	90	94

SOURCE: U.S. Department of Education, National Center for Education Statistics.

OMB Clearance Process and RELs

In order to examine the extent which OMB clearance processes affected the timeliness of randomized controlled trial design studies (RCTs) being conducted by the NCEE's RELs, OMB clearance process data were examined. This analysis was conducted at the request of some IES staff who expressed concerns regarding the perceived impact of the OMB clearance process on the timeliness of RCT research being conducted by the RELs. For the 26 RCTs that have completed the OMB clearance process during FY 06 and FY 07, data related to the dates for various steps in the process were examined. Table 16 notes the number of days that each request was at OMB before being approved, as well as the total number of days from start to finish for the clearance process (i.e., time from date entered into EDICS to OMB signoff). In general, it took an average of approximately 100 days at OMB and approximately 188 days total from start to finish.²⁹

²⁹ Three outliers were removed, two with much higher than average number of days (i.e., Southeast 2.1 and Southwest 1.1) and one with much fewer than average number of days (i.e., Central 1.1.). The means with all cases included are approximately 116 days at OMB and 203 days from start to finish.

Relevance

Table 16.—Length of time for Office of Management and Budget (OMB) clearance process for Regional Educational Laboratory (REL) projects using Randomized Controlled Trials (RCT)

REL projects	Number of days at OMB	Number of days from start to finish
Southeast		
Southeast 2.1	468	544
Southeast 1.1	117	196
Southeast 2.3	155	234
Southwest		
Southwest 1.1	237	321
Southwest 2.1	99	191
West		
West 2.5	114	274
West 2.4	110	208
West 2.6	110	208
West 2.?	113	189
West 2.?	115	189
Midwest		
Midwest 1.1	100	239
Midwest 2.1	99	216
Midwest 2.3	98	183
Mid-Atlantic		
Mid-Atlantic 2.2	106	203
Mid-Atlantic 2.1	87	160
Pacific		
Pacific 2.1	100	202
Central		
Central 2.1a	112	184
Central 1.2.9	84	162
Central 2.3c	80	156
Central 2d	76	153
Central 1.1	5	83
Northwest		
Northwest 2.2	105	178
Northwest 2.1	69	147
Northwest 1.1	70	140
Appalachia		
Appalachia 2.3	98	166
Northeast		
Northeast 2.3	86	156

SOURCE: U.S. Department of Education, Institute of Education Sciences.

NCER and NCSEER Grants

For the purposes of this evaluation, timeliness of NCER and NCSEER funded grants as it relates to relevance focuses primarily on the timeliness of the dissemination of findings from funded projects, particularly those studies employing randomized controlled trials to determine the efficacy or effectiveness of programs or interventions. In many ways, dissemination of findings for these grants is currently beyond the control of IES. Although proposals typically include a discussion of how findings and relevant research will be disseminated, these dissemination activities have traditionally been the responsibility of the principal investigator and often occur after the official time period for the grant has ended. Timeliness of dissemination of findings is also difficult to ascertain for NCER and NCSEER grants because the longitudinal nature of many of the funded projects, and the length of time it often takes to see impact on key outcome variables, means that many of the programs have not been funded long enough yet to be able to disseminate findings.

However, this evaluation did examine the earliest of the NCER programs focusing on randomized control trials: the Preschool Curriculum Evaluation Research Initiative. To provide needed evidence of the impact of contemporary preschool curricula, NCER conducted a multisite efficacy evaluation of 14 preschool curricula. In 2002, NCER awarded grants to seven researchers to implement several widely used preschool curricula, with Research Triangle Institute (RTI) International collecting common data across the seven projects. In 2003 NCER funded an additional five researchers, with Mathematica Policy Research (MPR), Inc. serving as their national evaluation coordinator. The final sample included Head Start, Title 1, State Pre-K and private preschool programs serving over 2,000 children in 20 geographic locations implementing 13 different experimental preschool curricula.

The evaluation of the preschool curricula occurred over 2 years, beginning with the preschool year in 2003–04 and continuing through the kindergarten year in 2004–05. Prekindergarten post-test data were collected in the spring from April to June 2004, and Kindergarten post-test data (student assessments, teacher reports, teacher surveys, and parent interviews) were collected in the spring and summer of 2005 between March and July. Therefore, final data collection occurred in July 2005. Findings from the multi-site evaluation were released July 2008 in the report, *Effects of Preschool curriculum programs on school readiness. Report from the Preschool Curriculum Evaluation Research Initiative*. This final report presents findings for the impact of each curriculum on five student-level outcomes (reading, phonological awareness, language, mathematics, and behavior) and six classroom-level outcomes (classroom quality, teacher-child interaction, and four types of instruction).

Relevance

The time from final data collection (July 2005) to the release of the final report (July 2008) was 3 years. Consequently, the field of education remained without rigorous research related to these preschool curricula for 3 years until the final report was released.

Summary of Findings

Is IES providing relevant, useful, and accessible data, research and publications to various stakeholder groups? To what extent does this relevance differ among stakeholder groups?

NCER: Given concerns about the validity and reliability of the GPRA data related to the relevance of NCER funded research from 2001 through 2006, it is not possible to draw any conclusions related to relevance of NCER projects or change over time. FY 07 baseline data for the new GPRA measure related to relevance of NCER projects were not available.

NCSER: For NCSER funded research, GPRA baseline data gathered in 2006 using an independent review panel indicated 50 percent of the funded NCSER research is highly relevant. Given the composition of the panel these assessments are likely to be more consistent and reliable measures over time.

NCES: According to the NCES customer survey, satisfaction with the relevance of NCES products, publications and services among core respondent groups (federal, state and local policymakers and academic researchers) have remained very high from 1997 through 2004. There were no apparent changes in satisfaction with relevance after the creation of IES, with satisfaction levels similar both before and after the implementation of IES.

NCES: According to the NCES customer survey, although still generally satisfied (76% satisfied or very satisfied), reporters and members of the media were the least satisfied of the various stakeholder groups with the relevance of NCES publications. This compares to a range of 91 percent to 98 percent for policymakers, supervisors/administrators/managers, teachers, researchers or evaluators, and other.

NCES: According to the NCES customer survey, although still generally satisfied (87% satisfied or very satisfied), policymakers were the least satisfied of the various stakeholder groups with the extent to which NCES information met their needs. This compares to an average of 94 percent for all other stakeholder groups. Policymakers were also least satisfied with the ease of obtaining the information, with 73 percent indicating they were satisfied or very satisfied as compared to an average of 88.5 percent for all other stakeholder groups.

The six interviewed stakeholders from major education-related organizations generally believed that IES should get “good marks” for relevance, although most persons also noted that they believed relevance has only more recently become a focus of IES. Some stakeholders also specifically noted a perceived difference in the relevance of research being funded by OERI versus IES, stating that IES appears to be better than OERI in its ability to tie research to the field.

To what extent is IES providing relevant data, and/or funding research and evaluation, that produce relevant findings as defined by the Institute's established priorities?

NCER/NCSER: The Institute appears to have effectively used its overall framework for its research grant programs (i.e., organizing programs within NCER and NCSER by outcomes such as reading or mathematics; type of education condition such as curriculum and instruction, teacher quality, administration, systems, and policy; grade level; and research goals) and its self-assessment process to identify gaps in existing research opportunities. IES has shown evidence of creating and modifying programs to ensure research opportunities meet the priorities.

NCER: The mix of funded NCER research grants for the most recent years does still resemble a triangle as identified by IES to be desirable (i.e., more identification and development activities, fewer small-scale field tests, and practices at scale at the apex). However, the base of identification and development grants is wider than the overall mix of grant applications noted in the 2006 NBES Annual Report as having the desirable triangular shape; and is also wider than the earlier years of funded research by the Institute. The average percentage of identification and development grants funded in 2006 through 2007 is 70.9 percent (N=56), representing a much broader base for the triangle as compared to 60 percent noted for grant applications in the 2006 NBES Annual Report; and the average percentage of efficacy grants funded in 2006 through 2007 is 2.5 percent (N=2), representing a much smaller apex of the triangle as compared to 9 percent.

NCER: The majority of content areas, with the exception of High School Reform and Education Policy, Finance and Systems are making virtually no use of the identification goal that focuses on identifying existing programs, practices, and policies that are differentially associated with student outcomes and the factors that mediate or moderate the effects of these programs, practices and policies.

NCER: There is an absence of scale-up research within the vast majority of content areas. For example, although the two Teacher Quality grant programs (i.e., Mathematics and Science Education and Reading and Writing) have funded a combined total of 37 grants, not a single scale-up grant has been awarded in either program. Between FY 02 and FY 07 a total of six scale-up grants have been awarded across all NCER content areas.

NCER: The relatively low numbers of funded NCER efficacy studies are particularly surprising in key, long-standing content areas such as Reading and Writing (seven efficacy studies between 2002 and 2007). The low numbers of efficacy studies in this key content area is not likely due to a lack of fully developed reading or mathematics programs in need of field tests of their efficacy. For example, in 2006–2007 the WWC examined 887 studies of 153 beginning reading programs. For beginning reading, 24 intervention programs met evidence standards, suggesting that the remaining 129 beginning reading programs identified

Relevance

by WWC represent a large population of programs needing to be rigorously examined via field-studies of efficacy. An additional six efficacy studies have been funded through the Teacher Quality: Reading and Writing Program. However, these grants have a slightly different focus and do not detract from the relatively low numbers of reading and writing efficacy studies that have been funded from 2002 through 2007.

NCSEER: The overall pattern of NCSEER funding for 2006 and 2007 is similar to NCER in terms of the majority of funded projects falling within the development category (i.e., 68% NCSEER versus 64.6% NCER funded between 2004 and 2007). In addition, for both NCER and NCSEER approximately 7 percent of funding was allocated for identification projects; and approximately 25 percent of funded grants fell within the other two aggregated categories for each Center (i.e., efficacy and scale-up).

There are significantly more proposals funded in the IES competitions that address student achievement outcomes than under OERI (nearly 25% increase). In addition there has been a steady increase in the percentage of IES NCER studies that have addressed student achievement outcomes (a 36.5% increase) from 2004 to 2007.

To what extent is IES producing findings and data in a timely manner that ensures their relevance to current and/or pressing education issues?

NCES: NCES has embedded within its infrastructure numerous measures of timeliness to ensure that the Center produces national databases within reasonable timeframes; and has focused its efforts since 2003 on reducing turnaround of database releases. For both 2006 and 2007, NCES met or exceeded its target in terms of timeliness of data releases. In addition, the percentage of NCES publications released within 18 months or less from the end of applicable data collection has increased each year, with a significant change from 2005 to 2006.

NCES: The 2004 NCES customer survey indicates that satisfaction with the timeliness of NCES databases has increased over time from 52 percent in 1997 to 78 percent in 2004. Satisfaction with the timeliness of NCES publications has ranged from 72 percent in 1997 to 77 percent in 1999 and 2004. Satisfaction with the timeliness of NCES services remained high across the survey years, averaging 89 percent. Pop-up, web-based surveys of a random sample of visitors to the NCES website during 2006-2007 also indicate high levels of satisfactions with NCES data files, publications and services (86% and 84% satisfied or very satisfied with timeliness of NCES data files in 2006 and 2007 respectively; 85% and 86% satisfied or very satisfied with timeliness of NCES publications in 2006 and 2007 respectively; and 92% and 94% satisfied or very satisfied with timeliness of NCES services in 2006 and 2007 respectively).

NCES: In terms of differences among stakeholder groups in satisfaction with timeliness, the 2004 NCES customer survey found that teachers expressed the highest level of satisfaction with the timeliness of NCES

publications (96%), and were second highest after supervisor, administrators and managers for timeliness of the release of NCES data (76% teachers, 79% supervisors, administrators and managers). For both NCES publications and NCES data, policymakers and reporters/members of the media reported the lowest levels of satisfaction (54% and 63% for publications respectively, and 61% and 49% for data respectively).

OMB and RELs: For 26 RCTs that have completed the OMB clearance process during FY 06 and FY 07, data related to the dates for various steps in the process were examined. Generally from the time the request was entered into EDICS to the date published on the Federal Register was less than 9 days (average of 8.3 days, with a range of 4 to 15). In general, it took an average of approximately 100 days at OMB and approximately 188 days total from start to finish.

NCER: For one of the earliest of the NCER programs focusing on randomized control trials, the Preschool Curriculum Evaluation Research Initiative in FY 02 and FY 03, the time from final data collection (July 2005) to the release of the final report (July 2008) was 3 years. Consequently, the field of education remained without rigorous research related to these preschool curricula for three years until the final report was released.

Page left intentionally blank.

UTILIZATION

A third explicit goal of IES is utilization, translating the results of education research into practice. As stated in the *IES 2005 Biennial Report to Congress*,

“Producing new education research that is both rigorous and relevant will help. However, the history of other fields suggests that more is involved in the use of good research than its mere presence. Evidence-based decisionmaking in other fields is enhanced by decision support tools that make the results of research available to users in easily understood forms. Education will adopt research-based approaches more rapidly if there are differential consequences for decisionmakers whose choices are not grounded in evidence, and it is easy to access and use such evidence.”

Therefore, given the explicit goal of IES in increasing utilization of rigorous research, the evaluation also addressed the following primary question: To what extent, and in which ways, has IES increased evidence-based decisionmaking (i.e., how is the rigorous and relevant research produced through the Institute’s efforts being used in education decisions)? In addition, in examining utilization, the evaluation also addressed the mechanisms for education decisionmaking. More specifically, the evaluation included the following question: How, and by whom, are education decisions related to policy and practice being made in the field?, and What are the implications for increasing the utilization of IES research, evaluation, publications, etc.?

The remainder of this section provides available data and findings related to each of these questions about the utilization of IES research and data, and the mechanisms for education decisionmaking. A brief summary of findings is also provided.

Utilization of IES Research and Data

To what extent, and in what ways, has IES increased evidence-based decisionmaking (i.e., to what extent is the rigorous and relevant research being used in education decisions)?

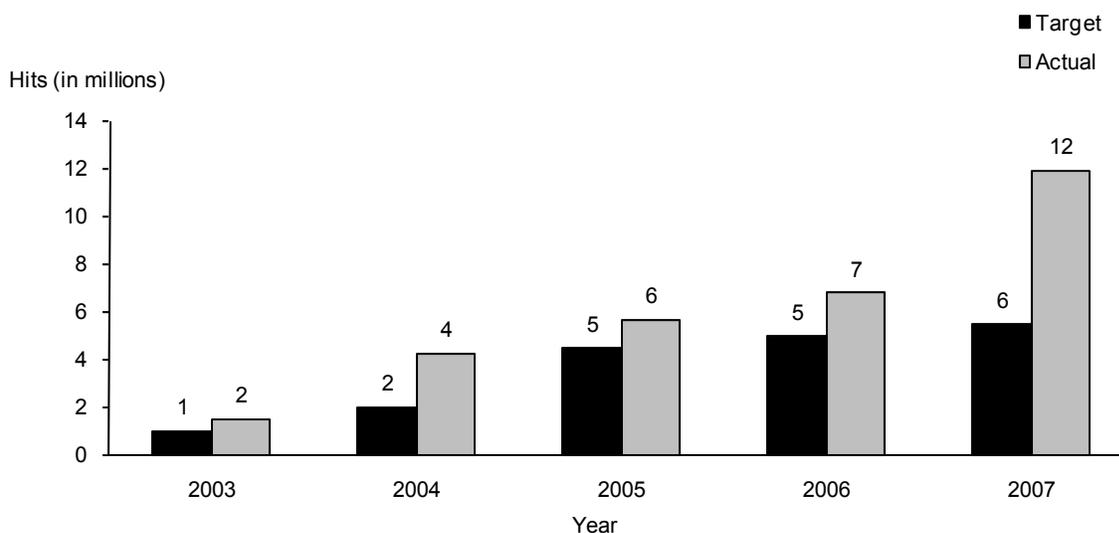
All accessible extant data related to the utilization of IES research, products and data were reviewed for the purposes of this evaluation. Available data primarily included data related to the NCEE products and services (i.e., WWC, ERIC, RELs) and data related to NCES products and services (NCES web-based hits, NCES product users, and external queries to NCES).

Utilization

NCEE

What Works Clearinghouse. Data related to the annual numbers of WWC hits were also examined for the purposes of the evaluation. The 2007 IES PART report measures the number of annual hits on the WWC website from 2003 to 2007.³⁰ Each year during this period the WWC exceeded its target, and annual hits increased from 1,522,922 in 2003 to 11,954,412 in 2007. Figure 15 provides the data on targets and page hits for each year; and table 17 provides details related to the actual numbers of hits.

Figure 15. What Works Clearinghouse annual website hits (in millions): 2003-2007



SOURCE: U.S. Department of Education, FY 2007 Program Performance Report.

Table 17. Number of annual hits on the What Works Clearinghouse website: 2003-07

Year	Target	Actual
2003	1,000,000	1,522,922
2004	2,000,000	4,249,668
2005	4,500,000	5,706,257
2006	5,000,000	6,794,141
2007	5,500,000	11,954,412

SOURCE: U.S. Department of Education, FY 07 Program Performance Report.

³⁰ The contractor for the WWC changed in 2007, and it is unclear whether this measure is continuing. WWC website usage data were obtained from IES for October 2007 through June 2008, however, these data do not appear to be equivalent to the PART data.

Data from a web-based pop-up survey on the WWC website provide some insight into the types of persons accessing the WWC data, and the stated purposes for visiting the WWC website. As noted in table 18, website visitors most frequently self-reported that they planned to use the information for either K-12 classroom or home instruction or curriculum development (22% each). Respondents less frequently noted planning to use the information obtained from the WWC for policy decisions: 11 percent each noted they planned to use the information for school or district policy decisions, 4 percent noted they planned to use the information for state policy decisions, and 3 percent stated they planned to use the information for federal policy decisions.

Table 18. WWC Website Survey: For what purpose do you plan to use the information you obtained from the What Works Clearinghouse website during this visit?

Purpose	Percent
K-12 classroom or home instruction	22
Curriculum development	22
Research project	13
School policy decision	11
District policy decision	11
State policy decision	4
Federal policy decision	3
Other	12

SOURCE: Institute of Education Sciences, WWC Website Survey.

The website pop-up survey also asked respondents the role or capacity in which they were currently visiting the WWC website. As noted in table 19 below, the survey indicates that teachers and administrators are the most frequent users of the WWC website (23% and 19% of all respondents, respectively). In addition, approximately 12 percent of respondents included researchers.

Utilization

Table 19. WWC Website Survey: In what capacity are you currently visiting the What Works website?

Capacity	Percent
Teacher (includes teachers and professors of all levels and types of education)	23
Administrator (principal, dean, department head, superintendent, etc.)	19
Researcher	12
School Support Staff (includes school guidance counselors, and paraprofessional schools personnel, including technology coordinators)	8
Local Education Agency (district)	4
Parent/Family (includes nuclear and extended family and child caregiver)	4
Program Developer/Vendor	4
State Education Agency	3
Student	3
Policymaker (board of education member; federal, state, or local public official; state or local education agency policymaker; legislator, etc.)	2
Technical Assistance Provider (includes staff of for- and non-profit education associations, Regional Educational Laboratories, and Professional Development Centers)	2
Other Federal Funds Recipient/Applicant (includes contractor, for- or non-profit organization, grantee, etc.)	2
Foundation Staff Member (includes personnel of organizations that fund grants and education venture capitalists)	1
Librarian (includes academic, federal, public, special, and state librarians and media specialists)	1
News/Media	1
Community Group Member (includes members of the business community, civic organizations, religious organizations, and volunteer groups)	1
Other	11

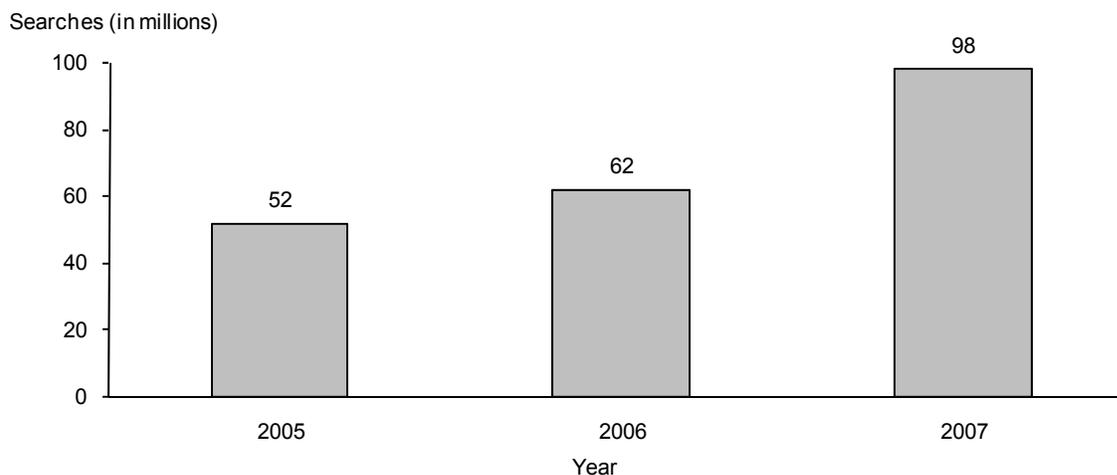
SOURCE: Institute of Education Sciences, WWC Website Survey.

Education Resources Information Center (ERIC) Usage. The Education Resources Information Center (ERIC) is the world’s largest digital library of education resources, with more than 1.2 million records and indexes of more than 600 journals; more than 80 percent of which are peer-reviewed. It provides access to bibliographic records of journal and non-journal literature from 1966 to the present. As such, ERIC has the capacity to serve as an important provider of rigorous education research for utilization by education decisionmakers.

Data regarding ERIC usage were examined, specifically the number of total ERIC searches using Google, vendors and the eric.ed.gov website and the average number of unique visitors per month using the

eric.ed.gov website.³¹ Figure 16 below indicates the estimated total ERIC searches for 2005 through 2007. As noted in the figure, total ERIC searches have increased over time, with a substantial increase occurring between 2006 and 2007.

Figure 16. Estimated number of ERIC searches (in millions): 2005-2007



SOURCE: U.S. Department of Education, Institute of Education Sciences.

In addition, data were available related to both the total ERIC searches and the average number of unique visitors per month using the eric.ed.gov site for two 3-month periods: October to December 2007 and January to March 2008. Data are provided in table 20 on the next page. Note that the data related to total searches for the 3-month period include ERIC searches using Google, vendors and the eric.ed.gov website; whereas the average number of unique visitors represents the number of persons per month during the three-month period and only includes searches conducted using eric.ed.gov.

³¹ Source: E-mail from Phoebe Cottingham to Norma Garza on 6/24/08, e-mail from Phoebe Cottingham to Steve Baldwin on 7/2/2008.

Utilization

Table 20. Total Education Resources Information Center (ERIC) searches and average number of unique visitors per month: 2007–08

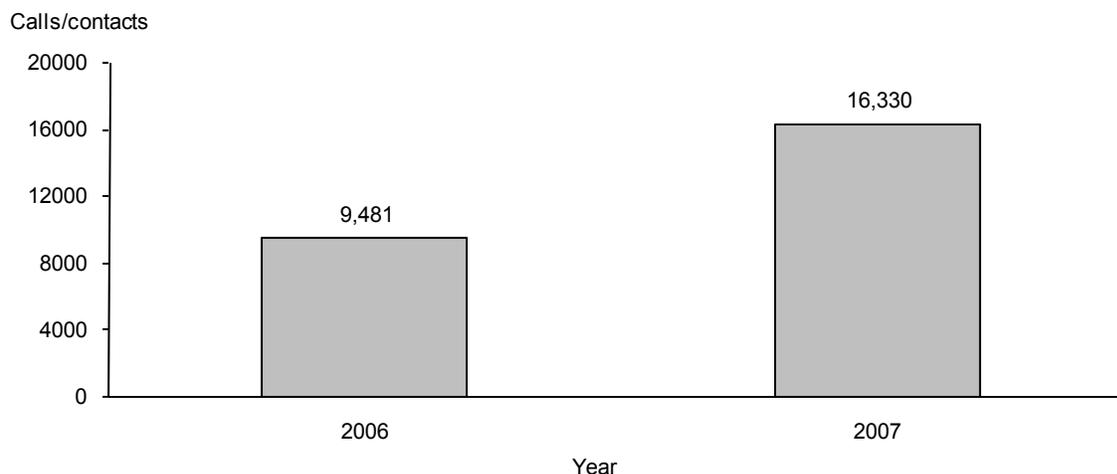
Time period	Total searches (Google, vendors, and eric.ed.gov site)	Average number of unique visitors per month using the eric.ed.gov site
October–December 2007	27,609,718	2,735,925
January–March 2008	28,776,644	2,977,775

SOURCE: U.S. Department of Education, Institute of Education Sciences.

These data indicate frequent usage of ERIC, including almost 56.4 million ERIC searches conducted within a 6-month period; and an average of more than 2.7 million unique visitors per month using the eric.ed.gov website to conduct ERIC searches. However, as with the WWC data reported previously, these data are also limited in terms of what they suggest regarding the use of IES data for evidence-based decisionmaking. There are no available data on which stakeholders are represented by these unique visitors, or on the purposes of the ERIC searches. These searches are likely to include graduate students and faculty using ERIC to conduct literature reviews and academic papers that may not be related to evidence-based decisionmaking in any way. Future indicators might address this by surveying ERIC users and collecting stakeholder data.

Regional Educational Laboratories. Data were provided related to the estimated number of calls and contacts received by the NCEE Regional Educational Laboratories (RELs) for 2006 and 2007.³² As depicted in figure 17 below, in 2007 there were 1.7 times more calls/contacts received than in 2006: 16,330 in 2007 versus 9,431 in 2006. The current priority for the 2006–2010 REL contract period is providing policymakers and practitioners with expert advice, training, and technical assistance on how to interpret the latest findings from scientifically valid research pertaining to requirements of No Child Left Behind. Given this priority, the increased number of calls and contacts may suggest an increased use of RELs for information needed in the education-decisionmaking process. However, additional details related to the types of stakeholders making these calls/contacts, and the basic purpose of the calls/contacts, would allow for the drawing of more reliable conclusions.

³² Source: E-mail from Morgan Stair to Steve Baldwin on 4/1/2008.

Figure 17. Estimated number of calls or contacts received by RELs: 2006-2007

SOURCE: U.S. Department of Education, Institute of Education Sciences.

NCES

Web-based hits. NCES provided data related to both page views and visits for the following websites for 2007: NCES and DAS.³³ Page views represent a hit to any file classified as a page. In order to view a web page with embedded images, for example, a browser must retrieve multiple files; and the page and its embedded files counts as a single page view. A visit is defined as a series of actions that begins when a visitor views his or her first page from the server, and ends when the visitor leaves the site or remains idle beyond the idle-time limit. Table 21 below provides the data related to both page views and visits for 2007. As noted in the table, the NCES website was visited frequently (i.e., 11.8 million visits per year), with the ratio of page views to hits of 6.5 to 1. DAS was accessed much less frequently, although there were still almost half a million visits per year.

³³ Source: E-mail from Jack Buckley to Steve Baldwin, sent 3/07/08.

Utilization

Table 21. National Center for Education Statistics (NCES) website statistics: 2007

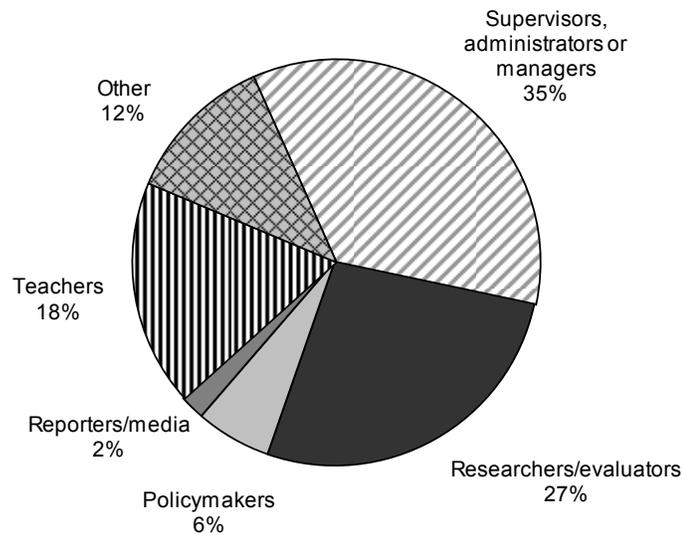
Website	Page views (in millions)	Visits (in millions)
NCES	76.6	11.80
Data Analysis System (DAS)	1.7	0.49

SOURCE: U.S. Department of Education, National Center for Education Statistics.

These data regarding NCES page views and visits are less limited in terms of what it suggests regarding the use of IES data for evidence-based decisionmaking given that other information related to NCES usage is available from the NCES customer surveys. Although the surveys do not focus exclusively on the website, the results do provide insight into what stakeholders are using NCES products; and also data regarding the stated purposes for using NCES and other education data. Findings from the customer survey related to which stakeholders use NCES products are presented in the following section.

NCES Product Users. Data from the 2004 NCES customer survey provide information on the types of stakeholders using NCES products. As noted in more detail in the section on relevance, the survey was administered to a random sample of over 3,900 federal policymakers, state policymakers, local policymakers, academic researchers, education association researchers, education journalists and known NCES users. NCES products were defined as “the NCES website, publications, web tools, [and] data files, excluding services such as responses to inquiries.” As indicated in figure 18, the largest percentage of NCES data users are supervisors, administrators or managers (35%) and researchers/evaluators (27%). Policymakers and reporters/media represented the smallest percentage of the distribution, with 6 percent and 2 percent respectively.

Figure 18. Distribution of NCES product users by involvement in education: 2004

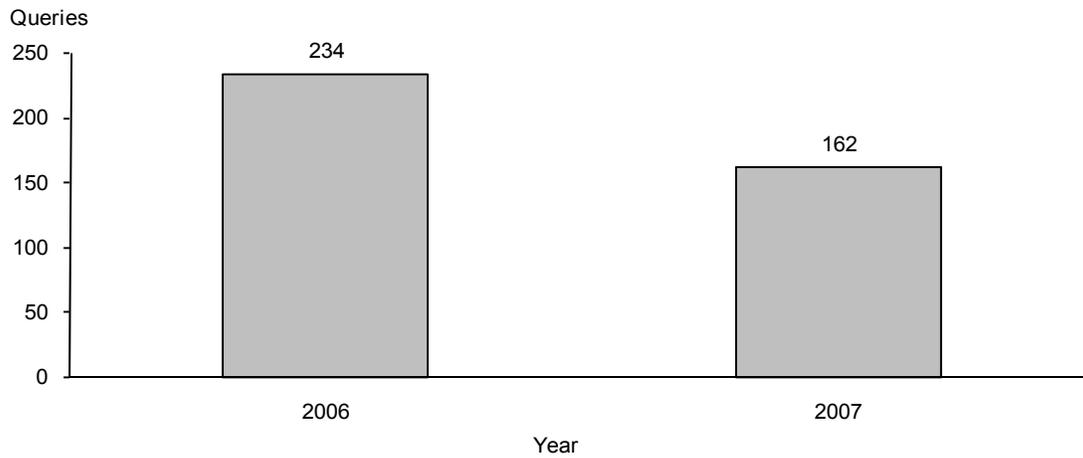


SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 Customer Satisfaction Survey.

External Queries to NCES. NCES internally tracks outside requests for data and requests for verification of data. Since October 2005 NCES has logged 541 such requests. Figure 19 provides the data for the 2 years with complete 12-month data: 2006 and 2007. As indicated in figure 19, requests to NCES appear to have declined over time, with 234 total in 2006 and 162 logged in 2007. In addition, for the first half of this year (i.e., January to July 2008) only 42 external queries were logged. For 2006 the volume of external queries from January to July was 149, and for 2007 there were 84 external queries from January to July.

Utilization

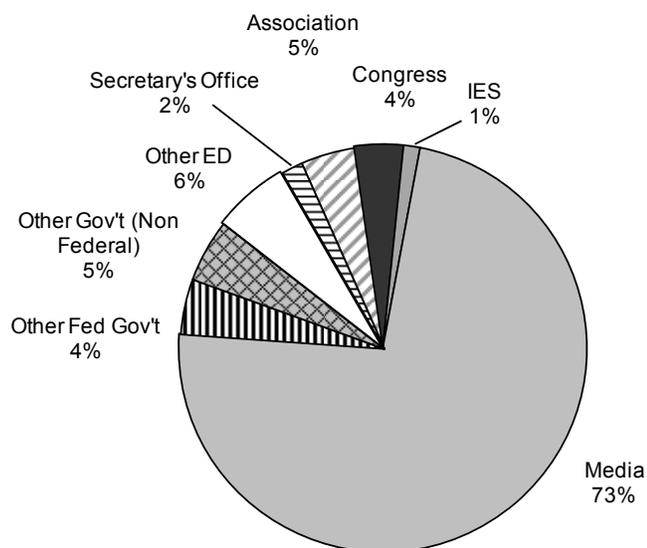
Figure 19. Number of external queries to NCES: 2006-2007



SOURCE: U.S. Department of Education, Institute of Education Sciences.

The two most popular surveys requested were the Common Core of Data (CCD), and the Integrated Postsecondary Education Data System (IPEDS) with approximately 17 percent and 16 percent of requests, respectively. The remainder of the requests (67% percent) were distributed among various other surveys and reports. NCES also tracks the source of the data request. Approximately 73 percent of all external queries from October 2005 to July 2008 were from the media. Figure 20 shows the distribution of external queries by the various stakeholder subgroups.

Figure 20. Distribution of external queries to NCES by organization type: 2006-2007



SOURCE: U.S. Department of Education, National Center for Education Statistics.

These data indicate that the decrease in overall external queries to NCES is the result of decreases in requests from the media that make the greatest numbers of requests. In other words, while requests from sources other than the media have remained relatively constant over the reporting period, the decline in total requests can largely be attributed to a decline in requests from the media. It is not clear why media requests have declined over time. For example, it could be that NCES data are becoming more accessible to the public and better utilized through websites and search engines, thus fewer direct requests to the center are needed. It could also be that the media have become less interested in NCES data, however, this would seem to contradict the general trend of increasing utilization of NCES data as shown by PART and website statistics. At the same time, it is unclear why requests from other organizations have not increased.

Mechanisms for Education Decisionmaking

How, and by whom, are education decisions related to policy and practice being made in the field? What are the implications for increasing the utilization of IES research, evaluation, publications, etc.?

General Education Information Needs

The 2004 NCES customer survey also gathered data related to the general education information needs of a random sample of more than 3,900 federal policymakers, state policymakers, local policymakers, academic researchers, education association researchers, education journalists, and known NCES users. Regardless of whether or not individuals used NCES products, all respondents were asked whether they used education data for any of a list of 13 purposes (Parker, Salvucci, and Wenk, 2005). The areas for which education data were used most frequently included the following: research or analysis (71%), general information (66%), and planning (60%). Data were least often used for marketing, sales or promotion (12%), updating databases (24%), and writing news articles or preparing material for media purposes (26%).³⁴

The 2004 NCES customer survey also asked both NCES product users, as well those who stated they had not used data from NCES in the past year, where they obtain education data other than NCES. Respondents could check as many of the 38 resources as were applicable, as well as reply to an open-ended “other” response. Education sources included federal government sources (e.g., other offices within the department of Education, Bureau of the Census, National Science Foundation), state and regional sources (e.g., state department of education, regional educational laboratories), national associations (e.g., American Council on Education, Council of the Great City Schools, American Federation of Teachers), and private research organizations and journals (e.g., Fordham Foundation, *Education Week*).

Table 22 notes the usage rates of the most frequently used non-NCES data sources by stakeholder group. Across all stakeholder groups, including both NCES users and nonusers, the top three most frequent non-NCES data sources consistently included the following two sources: “your state department of education” and “other offices within U.S. Department of Education.” For the following groups the U.S. Census Bureau was also amongst the top three: NCES-user policymakers, NCES-user researchers/evaluators, and for both user and nonuser reporters/media. Nonuser policymakers and both user and nonuser supervisors,

³⁴ Note that respondents could choose more than one area of education use.

Table 22. Percentage of National Center for Education Statistics (NCES) product users and non-users that report obtaining education data from various non-NCES sources, by stakeholder group: 2004

	Stakeholder group											
	All types		Policymaking		Supervision, administration, or management		Teaching		Research, evaluation, or testing		Reporting/ media	
	NCES product users	Non-users of NCES products	NCES product users	Non-users of NCES products	NCES product users	Non-users of NCES products	NCES product users	Non-users of NCES products	NCES product users	Non-users of NCES products	NCES product users	Non-users of NCES products
Non-NCES sources												
Any non-NCES source	98	87	100	95	98	93	98	93	98	82	99	89
Your state department of education	77	68	65	72	88	83	79	71	64	39	68	80
Other offices within U.S. Department of Education	73	49	66	62	75	57	80	59	66	33	74	44
U.S. Bureau of the Census	52	24	67	42	47	31	38	14	62	22	72	35
State or regional associations	48	36	61	47	65	50	41	27	31	16	51	33
<i>Education Week</i>	41	28	41	42	48	36	37	22	37	19	47	34
Educational Testing Service	36	20	52	36	40	26	24	13	40	18	41	25
State departments of education in other states	36	15	44	32	32	15	34	12	40	21	45	9
<i>Chronicle of Higher Education</i>	35	20	55	33	27	17	35	30	41	24	46	22
American Educational Research Association	34	22	26	33	24	12	50	45	43	39	15	12
Subject area association	34	24	26	30	36	24	49	42	23	14	28	29
U.S. Bureau of Labor Statistics	33	—	52	—	35	—	14	—	41	—	47	—
National Education Association	31	14	33	33	35	15	38	18	18	4	39	20
Association for Supervision & Curriculum Development	30	—	18	—	48	—	24	—	15	—	10	—
<i>Phi Delta Kappan</i>	29	—	—	—	39	—	29	—	24	—	10	—
National Science Foundation (independent agency of U.S. Government)	27	—	37	—	26	—	21	—	30	—	29	—

— Not available. Data was available only for those sources cited by at least 30 percent of all users/non-users or at least 30 percent of any involvement group. Therefore, although data are reported for NCES product users for the last three categories on the table, similar data are not included for NCES non-users because less than 30 percent of this population noted using these sources, and no single involvement group had higher than 30 percent for these sources.

NOTE: Respondents could select more than one data source. Table includes only sources cited by at least 30 percent of all users/non-users or at least 30 percent of any involvement group.

SOURCE: U.S. Department of Education, National Center for Education Statistics, 2004 Customer Satisfaction Survey.

Utilization

administrators or managers noted state or regional associations as one of the top three non-NCES data sources; and user and nonuser teachers as well as nonuser researchers/evaluators noted the American Educational Research Association as one of the top three most frequently used sources of education data. In general, across all stakeholder groups and sources of education data, NCES product users more frequently reported using each of the respective sources of information. The exceptions were as follows: policymaker nonusers reported slightly higher usage than policymaker users of NCES products for state departments of education, American Educational Research Association, and subject area association; and non-user reporters/media reported slightly higher usage of state departments of education than did reporters who used NCES data.

NCEE's Regional Educational Laboratories (RELs) were noted as a possible source of education data within the "state and regional sources" section. However, RELs do not appear in the chart above because fewer than 30 percent of all NCES data-users and fewer than 30 percent of all non-users noted that they obtained data from RELs; and fewer than 30 percent of any single stakeholder group within either users or nonusers obtained education data from RELs. Unfortunately, the WWC was not noted specifically as a possible education data source on the survey, and therefore the survey does not provide any data related to the frequency with which various stakeholders use (or do not use) the WWC. However, since this was not an explicit purpose of the NCES customer survey, the absence of the WWC from the resources is understandable.

Stakeholder Interviews

As noted previously, SEI/CEEP conducted key stakeholder interviews to supplement extant data. Interviews were completed with key stakeholders from the following organizations and associations: American Educational Research Association (AERA), American Psychological Association (APA), National Academy of Sciences, Council of the Great City Schools, Knowledge Alliance, and the National Sorority of Phi Delta Kappa. Interpretation is somewhat limited due to the small numbers of education-related organizations represented. However, the stakeholders included in the data represent some of the largest and most representative education-related organizations in the nation (e.g., AERA and APA); and the interview responses represent these persons' perceptions of the views and opinions of their broader constituencies, rather than the individual opinions of six persons. Therefore, the data from these six interviews do provide some valuable insight into perceptions of IES impact, particularly when interpreted within the context of other available data.

All stakeholders interviewed noted that there is an absence of information in the field about how education decisions are made within schools. In addition, these stakeholders frequently discussed their perception that there have been no changes in many years related to both the knowledge base of education decisionmaking and the mechanisms by which education decisions are made. Several representative comments include the following:

“The way decisions are made hasn’t changed since the [19]80s. The decisionmakers and leaders just don’t have the training to understand the importance of RCTs.”

“There is a level of sophistication needed to unearth and find and translate current research. It may be used by persons in schools of education and within the academy, but it’s not being used by practitioners in the field.”

“Not many have figured out how practitioners use research. Even if we learn that a certain intervention has an effect, there is a larger question about adherence and how we get practitioners to use it in the field.”

Given this perceived lack of knowledge in the field about the processes and mechanisms of education decisionmaking, most of the stakeholders tended to state that it was understandable that IES has not made much progress with regards to utilization. Some representative comments included the following:

“How to package and disseminate knowledge and information is so new, I don’t know if IES could have pushed it any further...I give IES lower marks in terms of utilization, not because they don’t want to make it happen or haven’t tried, we as a field just haven’t figured out yet what mechanisms we need to use to increase use by practitioners and policymakers.”

“The problems with utilization are not problems with IES, these problems are reflective of the field more generally...Everyone talks about translating research into practice, but no one knows how to do it.”

In terms of the three primary goals of IES (i.e., rigor, relevance and utilization), utilization uniformly received the lowest marks and the most criticism. However, one stakeholder stated that “both OERI and IES are bad at utilization, but at least IES seems to care more.”

Although questions related to specific IES centers or activities were not asked of interviewees, all stakeholders voluntarily noted that the WWC is the primary mechanism used by IES in its attempts to increase utilization. Some stated that the WWC was the “only real mechanism,” whereas other noted that utilization was also an intended purpose of the RELs. However, both the WWC and RELs were widely viewed as not being successful in increasing utilization of rigorous research. Unprompted comments related to the WWC included the following:

Utilization

“People are not going to the WWC because it basically says that we don’t know what works.”

“The WWC was overzealous in its evidentiary standards. No research was ever sufficient and it was almost impossible for any research to be respected. Only a very few studies met standards, and this left practitioners with nothing. Practitioners were in a bind because they were told they needed evidence-based practices, but at the same time were being told by WWC that nothings works. This optimal or nothing approach meant that practitioners stopped looking to the WWC for help.”

“WWC is focused on dissemination, but dissemination is not utilization. It’s trying to make research findings more accessible, but this is not going to get people to use the research.”

Almost all stakeholders also noted that RELs have not had an impact on increasing utilization of rigorous research. Unsolicited negative comments related to RELs were typically followed up with questions regarding potentially positive influences of the RELs. Most interviewees’ responses remained unchanged. However, one stakeholder quoted below did comment on the increased relevancy of the work of RELs. Comments related to the RELs and utilization included the following:

“RELs are supposed to be the translators of the research, but this part of the statute is not really implemented. If you look at their work plans there are three main areas: studies, rapid response research and technical assistance/help desks. But most of the focus is on the studies, which take 3 to 5 years, and not on translating studies and making them useful.”

“RELs are still trying to determine their role in the system. They are still trying to figure it out. They may be doing more relevant research than before, but they aren’t really having an impact on utilization.”

“Most superintendents do not know who their lab is or what they do or how to contact them. They know they [the RELs] exist and that’s about it.”

Many of the stakeholders noted the need for additional mechanisms by which to increase utilization. Some noted the need for additional studies of education decisionmaking to help inform what mechanisms would be most effective and efficient; and others noted the need for more rapid response mechanisms within the IES infrastructure that could provide practitioners with more timely information and guidance. Many also noted the need to move beyond what one stakeholder referred to as the “optimal or nothing” approach. These stakeholders noted the need to provide practitioners and education decisionmakers with guidance even if gold standards had not been met with any intervention research; and one stakeholder noted that “if two reports reach different conclusions, IES needs to take leadership in guiding people on how to interpret these for policy and practice.” One person also noted the need for more professional development and

technical assistance to make practitioners and policymakers better consumers of research. One stakeholder also recommended moving the role or function of utilization to another department or unit outside of IES.

Research on Evidence-Based Decisionmaking

An extensive review and synthesis of almost 30 years of research and literature conducted by Honig and Coburn (2008) found that the research base related to district-level administrators' use of evidence in decisionmaking is limited. However, the numerous studies that exist allowed the researchers to provide some information and insight into the forms of evidence used by central office administrators to make critical education decisions. Honig and Coburn found that most research suggests that district central office administrators use a broad range of evidence including “practitioner or local knowledge” that reflects information generated by practitioners or laypeople through their personal experience. In addition, they found that evidence does not directly inform decisions, but rather influences working knowledge which may in turn shape decisionmaking. The authors also found a number of factors that seem to shape central office use of evidence, including the following: features of the evidence itself, individual and collective working knowledge, social capital within and beyond the central office, district central office organization, institutional norms within district central offices, and political dynamics such as superintendent turnover. The authors found that “education policy including recent federal and state mandates on school district central offices to use evidence may affect evidence use but its influence appears to be mediated by these other factors” (pp. 594). Therefore, they conclude that these other forms of evidence may be necessary and critical to growing and sustaining the incorporation of evidence into day-to-day district central office decisions.

Honig and Coburn also discuss implications for policy and practice, as well as implications for education policy research. For example, Honig and Coburn state that “policymakers might advance evidence use if they acknowledged and provided specific supports for the subactivities fundamental to evidence use” (pp. 602). Examples include allocating time and resources for collaborative sensemaking processes that incorporating evidence seem to require; and professional development efforts aimed at preparing professionals across entire central offices (not just those in research and evaluation units) to use evidence in their decisionmaking. The authors also strongly note the need to build a stronger evidence base about which evidence administrators use, how they use it, and the conditions that help or hinder its use.

Summary of Findings

To what extent, and in what ways, has IES increased evidence-based decisionmaking (i.e., to what extent is the rigorous and relevant research being used in education decisions)?

WWC: The WWC exceeded its target for website hits each year from 2003 to 2007. Annual hits increased from 1,522,922 in 2003 to 11,954,412 in 2007.

WWC: Data from a web-based pop-up survey on the What Works Clearinghouse (WWC) website indicate that website visitors most frequently self-report that they plan to use the information for either K-12 classroom or home instruction or curriculum development (22% each). Respondents less frequently noted planning to use the information obtained from the WWC for policy decisions: 11 percent each noted they planned to use the information for school or district policy decisions, 4 percent noted they planned to use the information for state policy decisions, and 3 percent stated they planned to use the information for federal policy decisions.

WWC: Data from a web-based pop-up survey on the WWC website indicate that teachers and administrators are the most frequent users of the WWC website (23% and 19% of all respondents, respectively). In addition, approximately 12 percent of respondents included researchers.

ERIC: Data indicate frequent usage of ERIC, including almost 56.4 million ERIC searches conducted within a 6-month period; and an average of more than 2.7 million unique visitors per month using the eric.ed.gov website to conduct ERIC searches. Total ERIC searches have also increased over time, with a substantial increase occurring between 2006 and 2007.

WWC and ERIC: There were no accessible extant data that provide insight into who is accessing the WWC website or conducting ERIC searches and for what purposes. Future measures based on new data collection such as user surveys would help to determine the extent to which WWC or ERIC is being used for the purposes of education-decisionmaking.

RELS: Data for 2006 and 2007 indicate that there were 1.7 times more calls/contacts received in 2007 than in 2006. The increased number of calls and contacts suggests an increased use of RELs. Additional details related to the types of stakeholders making these calls/contacts, and the basic purpose of the calls/contacts, would help to draw more reliable conclusions on the relationship of these calls to education decisionmaking.

NCES: Website statistics for NCES indicate frequent visits (i.e., 11.8 million visits per year). The DAS has been accessed much less frequently, although almost half a million visits per year are reported. The NCES website clearly receives more hits than the WWC site: for NCES the average number of page views per month for 2007 was 6.38 million as compared to an average of 996,201 per month for the WWC in 2007.

NCES: The most likely users of NCES products and data are supervisors, administrators or managers (35%) and researchers/evaluators (27%). Policymakers and reporters/media represented the smallest percentage of the distribution, with 6 percent and 2 percent respectively.

NCES: External requests to NCES appear to have declined over time, with 234 total in 2006 and 162 logged in 2007. Data indicate that the decrease in overall external queries to NCES is the result of decreases in requests from the media, who make the greatest number of requests.

How, and by whom, are education decisions related to policy and practice being made in the field? What are the implications for increasing the utilization of IES research, evaluation, publications, etc.?

The 2004 NCES customer survey found that across all stakeholder groups, including both NCES users and non-users, the top three most frequent non-NCES data sources consistently included the following two sources: “your state department of education” and “other offices within U.S. Department of Education.” For the following groups the U.S. Census Bureau was also amongst the top three: NCES-user policymakers, NCES-user researchers/evaluators, and for both user and nonuser reporters/media. Nonuser policymakers and both user and nonuser supervisors, administrators or managers noted state or regional associations as one of the top three non-NCES data sources; and user and nonuser teachers as well as nonuser researchers/evaluators noted the American Educational Research Association as one of the top three most frequently used sources of education data.

The 2004 NCES customer survey indicates that fewer than 30 percent of all NCES data-users and fewer than 30 percent of all nonusers noted that they obtained data from RELs; and fewer than 30 percent of any single stakeholder group within either users or nonusers obtained education data from RELs.

Unfortunately, the WWC was not noted specifically as a possible education data source on the survey, and therefore the survey does not provide any data related to the frequency with which various stakeholders use (or do not use) the WWC. However, since this was not an explicit purpose of the NCES customer survey, the absence of the WWC from the resources is understandable.

In terms of the three primary goals of IES (i.e., rigor, relevance and utilization), utilization uniformly received the lowest marks and the most criticism from the six interviewed stakeholders from major education-related organizations. However, given a perceived lack of knowledge in the field about the process and mechanisms of education decisionmaking, the vast majority of stakeholders also tended to state that it was understandable that IES has not made much progress with regards to utilization.

The WWC was noted by these six stakeholders as the primary mechanism used by IES in its attempt to increase utilization. Some stated that the WWC was the “only real mechanism,” whereas other noted that utilization was also an intended purpose of the RELs. However, both the WWC and RELs were widely

Utilization

viewed by the interviewed stakeholders as not being successful in increasing utilization of rigorous research.

These six stakeholders generally noted the need for additional mechanisms by which to increase utilization. Some noted the need for additional studies of education decisionmaking to help inform what mechanisms would be most effective and efficient; and others noted the need for more rapid response mechanisms within the IES infrastructure that could provide practitioners with more timely information and guidance. Many also noted the need to move beyond what one stakeholder referred to as the “optimal or nothing” approach.

An extensive review and synthesis of almost 30 years of research and literature conducted by Honig and Coburn (2008) found that the research base related to district-level administrators’ use of evidence in decisionmaking is limited. The authors strongly note the need to build a stronger evidence base about which evidence administrators use, how they use it, and the conditions that help or hinder its use.

Existing studies allowed Honig and Coburn (2008) to provide some information and insight into the forms of evidence used by central office administrators to make critical education decisions. The authors found that “education policy including recent federal and state mandates on school district central offices to use evidence may affect evidence use but its influence appears to be mediated by these other factors” (pp. 594). Therefore, they conclude that these other forms of evidence may be necessary and critical to growing and sustaining the incorporation of evidence into day-to-day district central office decisions.

DISCUSSION AND RECOMMENDATIONS

Most persons within the field of education would agree that since the creation of the Institute of Education Sciences there has been an increase in the quantity of RCTs being conducted within the field of education, as well as increased dialogue within the education research community regarding what constitutes rigorous research. An analysis of published journal articles by Conostas (2007) supports this general view. Comparing data for 2001 (prior to the establishment of IES) to data for 2005, Conostas found an increased use of terms representing federal priorities for education research (i.e., experimental, randomization, hypothesis, and quantitative). For example, Conostas found that the number of published journal articles that contained the term “random” in the title, the abstract, or a descriptor increased by 219 percent. Stakeholder interview data from this evaluation also supports this general view. Although somewhat limited in terms of interpretation due to the small numbers of education-related organizations represented, the stakeholders included in the data represent some of the largest and most representative education-related organizations (e.g., AERA and APA) and therefore do provide some valuable insight into perceptions of IES impact. There is a consensus among interviewed stakeholders that IES has played a major role in increasing dialogue within the education community related to what constitutes rigorous methodology; and a belief that the Institute has increased awareness and utilization of randomized controlled trial (RCT) design studies. Even those stakeholders who have commonly expressed strong criticism of the emphasis the Institute has placed on RCTs clearly stated that the Institute has increased the quality of research being conducted within the field of education; and stated that that the emphasis on rigor is much stronger and more pronounced within IES than it was during the era of OERI.

However, determining the impact of the Institute on rigor has the same problems that other education research issues face in attempting to establish causal relationships. The difficulty remains in separating out causation from correlation. Unfortunately, causal claims regarding the impact of the Institute on rigor, relevance and utilization cannot be made within the scope of this evaluation, or perhaps within the scope of any feasible evaluation. Clearly the Elementary and Secondary Education Act (ESEA) of 2001 and the Education Sciences Reform Act of 2002 (ESRA), as well as the strong accountability standards included as part of No Child Left Behind, have also contributed to changes related to increasing emphasis on rigor and scientific standards. Although causation cannot be determined, there are several general conclusions related to the Institute’s focus on rigor and RCTs that can be made from the extant data accessible for this evaluation.

Discussion and Recommendations

First, the emphasis and attention to rigorous methodology is clearly more prominent within the Institute than it was within its predecessor, OERI. Clear examples of the focus IES has placed on RCTs and rigorous methodology are evident from the structure used for NCER grant programs that includes two goals focused specifically on using rigorous methodology (especially RCTs) to measure efficacy and effectiveness, and the focus and attention placed on the What Works Clearinghouse. In addition, the fact that demand has exceeded capacity for the summer institutes on cluster randomized trials for both 2007 and 2008 indicate that education researchers understand the importance of RCTs in the funding priorities of the Institute.

Second, there has been a sharp increase in the number of RCTs being conducted within IES as compared to OERI. For example, whereas 32 percent of funded projects addressing causal questions used RCTs just prior to the establishment of IES in 2001, 82 percent to 100 percent of NCER new research and evaluation projects addressing causal questions used RCTs in the years following the establishment of IES. In addition, 24 large IES-supported evaluation studies using rigorous methodology are currently underway as opposed to just one such evaluation study in 2000 under the support of OERI.

Third, analysis of NCER and NCSER efficacy and effectiveness funded proposals for FY 04 through 2007 on 10 dimensions of high quality research designs suggests that these IES studies have a high potential for generating rigorous and valid evidence of effectiveness. Although accessible extant data is not yet available for the vast majority of these studies, analyses indicate that over time increasing percentages of funded efficacy and effectiveness proposals have included these dimensions of high quality research. However, the extent to which these designs are being implemented with fidelity cannot yet be determined.

Finally, IES has placed a strong emphasis on increasing the capacity of the field to conduct rigorous research. To date NCER has funded 242 predoctoral fellows (2004 through 2008) and 30 postdoctoral fellows (2005 through 2008); and in July 2008 NCSER awarded five new grants for postdoctoral special education training fellowships. In addition, IES has recently begun implementing training institutes and seminars to increase researchers' skills and capacity in conducting rigorous education research (i.e., cluster randomized trials, evaluating state and district level interventions and single-case design). Demand has exceeded capacity for the 2-week intensive summer institute trainings on cluster randomized designs for both 2007 and 2008, suggesting that there is substantial interest from the field in increasing capacity related to rigorous methodology. What remains unknown regarding these IES initiatives is the extent to which they are effective in increasing the quantity and quality of rigorous education research. For example, although preliminary data indicates that 80 percent of the persons who have completed their predoctoral fellowships are employed in research positions of some type, what remains unknown is the extent to which these

interdisciplinary fellows actually pursue a research agenda related to education, and the extent to which these fellows will contribute rigorous research to the field of education.

In terms of the three primary goals of the Institute of Education Sciences (i.e., rigor, relevance and utilization), the Institute has clearly made the most visible and prominent contribution within the area of rigor. Stakeholders interviewed for the evaluation generally believed that IES should get “good marks” in relevance, but also stated that they believed relevance has only more recently become a focus of IES. In terms of relevance, there is little reliable or valid data that provide insight into possible changes over time. The most current GPRA data suggests that substantial work still needs to be done in increasing the relevance of NCER and NCSER funded research: independent, external review panels found that 50 percent of funded NCSER research and 33 percent of funded NCER research is highly relevant. NCES has also historically collected data related to relevance through its customer survey. Findings generally indicate high levels of satisfaction with the relevance of NCES products, publications and services from 1997 through 2004, with levels of satisfaction similar both before and after the implementation of IES. NCES also examined differences in relevance amongst stakeholder groups, finding that although still generally very satisfied, reporters were the least satisfied with the relevance of NCES publications, and policymakers were least satisfied with the ease of obtaining information from NCES.

Relevance within the Institute was also examined in terms of the extent to which NCER and NCSER funding was aligned with the goals and priorities established by IES. In general, the Institute appears to have effectively used its overall framework for its research grant programs and its self-assessment process to identify gaps in the existing research opportunities, and has shown evidence of creating and modifying programs as needed. However, given that the most relevant and practical evidence from the perspective of practitioners and policymakers is likely to come from efficacy and effectiveness research, the absence of scale-up research within the vast majority of content areas (e.g. although the two Teacher Quality grant programs, Mathematics and Science Education and Reading and Writing, have funded a combined total of 37 grants, not a single scale-up grant has been awarded in either program; and between FY 02 and FY 07 a total of six scale-up grants have been awarded across all NCER content areas) raises some concerns in terms of relevance of the research and findings to the field. In addition, the relatively low numbers of efficacy studies in some key, long-standing content areas with relatively large research bases such as such as Reading and Writing (seven efficacy studies between 2002 and 2007) are somewhat surprising. Regardless of whether or not this is an issue of a lack of capacity amongst education researchers to conduct this type of research, as suggested by IES, there are clear implications for the relevance of the research to the field.

Discussion and Recommendations

Timeliness is also a factor in considering the relevance of findings and data. It is clear that NCES has embedded within its infrastructure numerous measures of timeliness, and has successfully focused its efforts on reducing turnaround time for both database releases and publications. However, a specific focus and emphasis on timeliness was not evident in the data available from the other Centers. Data related to NCER's Preschool Curriculum Evaluation Research (PCER) Initiative raises concerns about the timeliness of findings related to rigorous research. The time from final data collection for these FY 02 and FY 03 programs to the release of the final report (and individual project findings) was 3 years, with the published final report released July 2008. Given that most other programs began too recently to have final data and reports, as well as the fact that most other NCER and NCSER content areas do not include a comprehensive external evaluation component like PCER, this timeliness issue may be an anomaly. The next few years will make it more apparent whether or not the lack of timeliness was specific to the PCER program, or indicative of a broader issue with NCER funded research.

Similar to relevance, stakeholders interviewed also generally agreed that utilization was not as strong of a focus for the Institute as was rigor. In fact, in terms of the three primary goals of IES (rigor, relevance and utilization), utilization uniformly received the lowest marks and most criticism from interviewed stakeholders. Valid and reliable data to confirm or disconfirm these stakeholder perceptions are not available. Data related to ERIC usage and REL calls/contacts are limited in their meaningfulness given the lack of information about who is accessing these sites/resources and for what purposes. The 2004 NCES customer survey does provide some insight into the types of data being used by various stakeholder groups. However, there is a general absence of knowledge and understanding within the field of education research about how to increase utilization of rigorous research by practitioners and policymakers. Given this lack of understanding and knowledge, it is understandable that IES has focused primarily on increasing dissemination of information. However, without a better understanding of the ways in which rigorous research can best be integrated into policy decisions and education decisionmaking, it will be difficult for the Institute to move beyond simply increasing dissemination efforts to truly increasing utilization.

In addition to using the accessible extant data to generate these findings related to the impact of IES on rigor, relevance and utilization, the evaluation also focused on developing recommendations related to evaluating IES impact, as well as broader recommendations regarding the priorities and practices of the Institute. These recommendations include the following:

Indicators/Performance Measures. IES's research, development, and dissemination programs recently received an *effective* rating, the highest score, on OMB's Program Assessment Rating Tool (PART). Given that the effective rating has only been given to 18 percent of more than 1,000 programs assessed by OMB, it is clear that the Institute has established generally strong indicators and performance measures for its programs and activities. However, there are still ways in which the indicators and performance measures can be modified, or new measures developed, to further strengthen the Institutes' ability to measure the impact of the Institute on rigor, relevance and utilization. For example, the current GPRA indicator based on the percentage of NCER funded research projects that are deemed to be of high quality is questionable in terms of reliability and validity, and a measure that is independent of the funding process itself would be more meaningful. Returning to a method of having an independent panel of experts reviewing funded proposals (such as in the 2002 GPRA data) removes the assessment of quality from the funding mechanism. Such a review, could serve as an external check on the reliability of the scientific review process associated with funding decisions. Related to the review process, having systematic criteria for what constitutes high quality in the four domains of significance, research plan, personnel and resources could make these two review processes even more comparable. To establish reliable trend data, this external review panel could conceivably rate samples of projects funded from OERI as well, given the only mode of comparison for impact is that of time (with baseline measures during OERI funding years, and follow up during IES-funded years).

Additional areas for improvement regarding specific indicators are evident throughout the evaluation report, including the following:

- (a) For the relevance GPRA indicator, the external panel used to rate the relevance of NCER funded projects should include representatives of national educational associations (similar to the panel for NCESER) that can provide broader input than the individual principals and superintendents currently on the external review panels.
- (b) Relevance indicators should include policymakers to help provide a measure of relevance to this stakeholder group, and/or the measure should specify that it pertains specifically to relevance for practitioners.
- (c) To increase the reliability and consistency of relevance and quality measures over time, external review panels need to remain relatively stable in composition over time, and clearly delineated rubrics and standards for rating need to be established. In addition, measures of inter-rater reliability and reliability of ratings over time should be included.

Discussion and Recommendations

(d) Indicators related to the pre- and postdoctoral training programs need to specifically address the extent to which these individuals' postfellowship employment is specifically related to research in education, rather than simply engaged in research, particularly given the interdisciplinary nature of the fellowships. Given the resources invested into these programs, it would also be useful to collect longitudinal data related to the area of employment/research and research productivity of pre-and postdoctoral fellows. Similar data gathered from participants in the intensive summer training institutes (e.g., quantity and quality of rigorous educational research conducted prior to training and postinstitute) might also provide useful comparative data related to the efficiency and effectiveness of these two mechanisms for increasing the capacity of the field to conduct rigorous research.

(e) Although PART assessment data includes gathering data in 2012–2014 on the percentage of persons who consult the WWC prior to making a decision, it would be helpful to also gather such data now to provide a better understanding of the extent to which the usage of WWC changes over time.

(f) Similar to the surveys that have historically been conducted by NCES, it would be useful to also periodically collect data from a representative sample of key stakeholders (e.g., practitioners, administrators, state and federal policymakers) regarding perceptions of quality and relevance, as well as behaviors related to utilization. Unlike data obtained from web-based pop-up surveys that only gather data from those persons already using IES products or services, this type of systematic survey would provide meaningful formative and summative data related to impact on rigor, relevance and utilization.

(g) Gathering systematic performance measure data from NCER and NCSER grantees would provide a more comprehensive and consistent measure of the quality, timeliness, relevance and utilization of the data and findings generated by these grants. Systematic data can be provided by each of the grantees; and final products could also be reviewed and rated for the quality and rigor of study implementation and findings.

(h) Data for calls/contacts received by RELs should be augmented by information on the types of inquirers and the purpose of their calls/contacts in order to provide a better understanding of the utilization of REL resources and services.

(i) WWC users/stakeholders should be surveyed about the relevance and utility of intervention reports, topic reports, quick review documents and practice guides in order to provide a better understanding of the utilization of these products and their role in education-decisionmaking.

(j) A specific focus on timeliness similar to that of NCES should be implemented by NCER and NCSER to ensure that findings from funded grants are disseminated in a timely manner.

NCER and NCSER Research Grant Findings. Currently, systematic extant data related to findings from funded NCER and NCSER projects are very limited. The lack of data makes it difficult to assess the rigor and relevance of research findings generated from these funded projects. Annual performance reports, while

using standardized forms, do not always yield the kind of information that can be used to assess the level of rigor associated with the research study; and assessing the quality of proposals can only provide a proximate estimate of what the level of rigor might be if the proposed methodology is implemented as originally planned. And the vast majority of projects do not have easily accessible reports or data available via the IES website. In addition to making it difficult to assess the rigor of completed studies, the lack of systematic extant data related to findings from NCER and NCSER funded projects also decreases the accessibility of these research findings, and therefore detracts from the possible utilization of the research findings by researchers, practitioners and policymakers. To increase the likelihood of utilization, as well as increase the ability to assess rigor of methodology as implemented, IES should consider making project reports more readily accessible to the public, as well as perhaps creating mechanisms for the systematic collection of data (e.g., align reporting requirements for efficacy and effectiveness studies to meet the standards of evidence criteria set out by the What Works Clearinghouse and provide a venue for detailing changes to the proposed methodology).

Capacity of Field to Conduct Rigorous Research. Given the strong interest expressed in the intensive summer training institutes on cluster randomized trials (i.e., demand exceeded capacity) and other methodological trainings, consideration should be given to expanding these programs. Since these intensive trainings target persons already in the field of education conducting research, and persons with strong interest in applying rigorous methodology to education settings, there seems to be the potential for substantial impact with relatively minimal costs compared to programs such as the predoctoral training program. Although the impact of the predoctoral fellowship program will not be evident for at least several years given the length of time needed for these individuals to begin contributing to rigorous research in education, the relatively high costs per student are readily apparent. For example, analyses of available data indicate that the average expenditure per student by predoctoral program is approximately \$176,000, with a range of approximately \$92,000 to \$333,000 per predoctoral fellow. Current estimates indicate a maximum of 80 percent of these predoctoral fellows conduct research postfellowship, and because the programs are interdisciplinary it is possible many of these fellows will not directly contribute to education research. The costs of the predoctoral fellowships do not indicate that these fellowships are not productive or imply that they should not be continued. Further data related to impact are still needed. But the cost data does suggest that further thought should be given as to whether or not there are other mechanisms that may more quickly and efficiently increase the capacity of the field to conduct education research, such as the intensive summer training institutes. For any alternative mechanisms for increasing capacity it will be important to develop and implement measures to examine the impact of these endeavors (e.g., number of participants who successfully receive IES funding for cluster randomized trials), as well as conducting cost-benefit

Discussion and Recommendations

analyses comparing the various mechanisms for increasing capacity of the field to conduct rigorous research.

Utilization. There is a clear and definite need in the field of education for a stronger research base related knowledge use (i.e., how to increase policymakers and practitioners use of rigorous research for education decisionmaking). There is little information currently available regarding the types of evidence practitioners, administrators and policymakers use, how they use it, and what conditions help or hinder its use. Without this knowledge, IES is likely to continue to focus on increasing access to rigorous research and the dissemination of rigorous evidence rather than employing strategies that truly increase utilization of rigorous research. Although access and dissemination are critical aspects of utilization, the research base on knowledge utilization that does exist suggests that the impact of these activities will remain minimal without a stronger understanding of knowledge utilization.

The complexities of increasing utilization are acknowledged in the IES PART long-term outcome measure that focuses on the percentage of decisionmakers surveyed in 2013–2014 who indicate they consult the What Works Clearinghouse prior to making decision(s) on reading, writing, math, science or teacher quality interventions. The target set for 2013–2014 is 25 percent, noted by IES in the PART document to be an ambitious goal. In other words, the long-term goal for the primary IES mechanism for increasing utilization is only 25 percent. Granted, IES is probably correct that this goal of 25 percent utilization is ambitious given that the research base on knowledge utilization suggests that policymakers and practitioners do not simply access available data and use these data to make education decisions. This type of linear relationship between rigorous evidence and decisionmaking does not exist. Therefore, a clear and strong research agenda related to better understanding how to increase the utilization of rigorous research among education practitioners and policymakers is needed. Without such a knowledge base the resources used to increase the rigor of education research will largely remain wasted as the rigorous research that produces findings regarding “what works” will only minimally be used in education practice or policy.

Future Evaluations. Appropriate resources, and latitude in terms of scope of work, need to be given to any future evaluations aimed at assessing the extent to which the Institute has been effective in carrying out its priorities and mission. The validity and meaningfulness of findings related to the impact of IES are substantially limited when only extant data can be used for the purposes of the evaluation. There are many meaningful and useful analyses that could be included as part of an evaluation of IES if additional resources and original data collection was allowed. For example, to measure the quality and relevance of NCER funded research over time a random sample of projects from each year during both OERI and IES could be

selected, and subsequently subjected to blind reviews (i.e., no information on the year of the proposal) by an appropriate panel of experts using carefully constructed scoring rubrics. Also, the evaluation of the impact of IES on rigor, relevance, and utilization could be enhanced by including surveys and/or interviews with past and current NCER and NCSER grantees. Data gathered through such surveys and interviews would provide the types of data needed to more validly measure the rigor and relevance of grants, and provide needed data not currently available through IES. Surveying and/or interviewing NCER and NCSER panel reviewers would be another possible method that would provide needed data to address key evaluation questions. The requirement to use extant data for this evaluation necessitated a backward mapping process whereby accessible extant data sources defined (and limited) the evaluation questions that could be addressed. Future evaluations of the effectiveness of IES in carrying out its mission need to allow the key evaluation questions to drive the design and methodology of the study.

Page left intentionally blank.

REFERENCES

- Constas, M.A. (2007). Reshaping the methodological identity of Education Research: Early signs of the impact of federal policy. *Evaluation Review*, 31(4): 391-400.
- Hoffer, T.B., Welch, V., Jr., Webber, K., Williams, K., Lisek, B., Hess, M., Loew, D., and Guzman-Barron, I. (2006). *Doctorate Recipients RFrom United States Universities: Summary Report 2005*. Chicago: National Opinion Research Center. (The report gives the results of data collected in the Survey of Earned Doctorates, conducted for six federal agencies, NSF, NIH, USED, NEH, USDA, and NASA by NORC.)
- Honig, M.I., and Coburn, C. (2008). Evidence-based decisionmaking in school district central offices: Toward a policy and research agenda. *Education Policy*, 22(4): 578-608.
- National Board for Education Sciences. (July 2006). *Annual Report for IES*.
- National Research Council. (1999). *Improving Student Learning: A Strategic Plan for Education Research and Its Utilization*. Washington, DC: National Academies Press.
- Parker, A.C.E., Salvucci, S., and Wenck, S.R. (2005). *2004 NCES Customer Satisfaction Survey Report* (NCES 2005-602). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Preschool Curriculum Evaluation Research Consortium. (2008). *Effects of Preschool Curriculum Programs on School Readiness* (NCER 2008-2009). National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Washington, DC.
- Walker, G. (2008). Admission requirements for education doctoral programs at top 20 American universities. *College Student Journal*, 42(2): 357-366.
- Whitehurst, G.J. (2005). *Biennial Report to Congress: Institute of Education Sciences*. Retrieved on July 20, 2008, from <http://ies.ed.gov/pdf/biennialrpt05.pdf>.