

**DEPARTMENT OF EDUCATION**

RIN 1890-ZA00

**Scientifically Based Evaluation Methods****AGENCY:** Department of Education.**ACTION:** Notice of final priority.

**SUMMARY:** The Secretary of Education announces a priority that may be used for any appropriate programs in the Department of Education (Department) in FY 2005 and in later years. We take this action to focus Federal financial assistance on expanding the number of programs and projects Department-wide that are evaluated under rigorous scientifically based research methods in accordance with the Elementary and Secondary Education Act of 1965 (ESEA), as reauthorized by the No Child Left Behind Act of 2001 (NCLB). The definition of scientifically based research in section 9201(37) of NCLB includes other research designs in addition to the random assignment and quasi-experimental designs that are the subject of this priority. However, the Secretary considers random assignment and quasi-experimental designs to be the most rigorous methods to address the question of project effectiveness. While this action is of particular importance for programs authorized by NCLB, it is also an important tool for other programs and, for this reason, is being established for all Department programs. Establishing the priority on a Department-wide basis will permit any office to use the priority for a program for which it is appropriate.

**EFFECTIVE DATE:** This priority is effective February 24, 2005.

**FOR FURTHER INFORMATION CONTACT:** Margo K. Anderson, U.S. Department of Education, 400 Maryland Avenue, SW., room 4W333, Washington, DC 20202-5910. Telephone: (202) 205-3010.

If you use a telecommunications device for the deaf (TDD), you may call the Federal Relay Service (FRS) at 1-800-877-8339.

Individuals with disabilities may obtain this document in an alternative format (e.g., Braille, large print, audiotape, or computer diskette) on request to the contact person listed under **FOR FURTHER INFORMATION CONTACT**.

**SUPPLEMENTARY INFORMATION:****General**

The ESEA as reauthorized by the NCLB uses the term *scientifically based research* more than 100 times in the context of evaluating programs to determine what works in education or

ensuring that Federal funds are used to support activities and services that work. This final priority is intended to ensure that appropriate federally funded projects are evaluated using scientifically based research. Establishing this priority makes it possible for any office in the Department to encourage or to require appropriate projects to use scientifically based evaluation strategies to determine the effectiveness of a project intervention.

We published a notice of proposed priority in the **Federal Register** on November 4, 2003 (68 FR 62445). Except for a technical change to correct an error in the language of the priority, one minor clarifying change, and the addition of a definitions section, there are no differences between the notice of proposed priority and this notice of final priority. The definitions section provides the generally accepted meaning for technical terms used throughout the document.

**Analysis of Comments**

In response to our invitation in the notice of proposed priority, almost 300 parties submitted comments on the proposed priority. Although we received substantive comments, we determined that the comments did not warrant changes. However, we have reviewed the notice since its publication and have made a change based on that review. An analysis of the comments and changes is published as an appendix to this notice.

**Note:** This notice does not solicit applications. In any year in which we choose to use this priority, we invite applications for new awards under the applicable program through a notice in the **Federal Register**. When inviting applications we designate the priority as absolute, competitive preference, or invitational. The effect of each type of priority follows:

**Absolute priority:** Under an absolute priority we consider only applications that meet the priority (34 CFR 75.105(c)(3)).

**Competitive preference priority:** Under a competitive preference priority we give competitive preference to an application by either (1) awarding additional points, depending on how well or the extent to which the application meets the competitive preference priority (34 CFR 75.105(c)(2)(i)); or (2) selecting an application that meets the competitive priority over an application of comparable merit that does not meet the priority (34 CFR 75.105(c)(2)(ii)).

When using the priority to give competitive preference to an application, the Secretary will review applications using a two-stage process. In the first stage, the application will be reviewed without taking the priority into account. In the second stage of

review, the applications rated highest in stage one will be reviewed for competitive preference.

**Invitational priority:** Under an invitational priority we are particularly interested in applications that meet the invitational priority. However, we do not give an application that meets the invitational priority a competitive or absolute preference over other applications (34 CFR 75.105(c)(1)).

**Priority**

The Secretary establishes a priority for projects proposing an evaluation plan that is based on rigorous scientifically based research methods to assess the effectiveness of a particular intervention. The Secretary intends that this priority will allow program participants and the Department to determine whether the project produces meaningful effects on student achievement or teacher performance.

Evaluation methods using an experimental design are best for determining project effectiveness. Thus, when feasible, the project must use an experimental design under which participants—e.g., students, teachers, classrooms, or schools—are randomly assigned to participate in the project activities being evaluated or to a control group that does not participate in the project activities being evaluated.

If random assignment is not feasible, the project may use a quasi-experimental design with carefully matched comparison conditions. This alternative design attempts to approximate a randomly assigned control group by matching participants—e.g., students, teachers, classrooms, or schools—with non-participants having similar pre-program characteristics.

In cases where random assignment is not possible and participation in the intervention is determined by a specified cutting point on a quantified continuum of scores, regression discontinuity designs may be employed.

For projects that are focused on special populations in which sufficient numbers of participants are not available to support random assignment or matched comparison group designs, single-subject designs such as multiple baseline or treatment-reversal or interrupted time series that are capable of demonstrating causal relationships can be employed.

Proposed evaluation strategies that use neither experimental designs with random assignment nor quasi-experimental designs using a matched comparison group nor regression discontinuity designs will not be considered responsive to the priority

when sufficient numbers of participants are available to support these designs. Evaluation strategies that involve too small a number of participants to support group designs must be capable of demonstrating the causal effects of an intervention or program on those participants.

The proposed evaluation plan must describe how the project evaluator will collect—before the project intervention commences and after it ends—valid and reliable data that measure the impact of participation in the program or in the comparison group.

If the priority is used as a competitive preference priority, points awarded under this priority will be determined by the quality of the proposed evaluation method. In determining the quality of the evaluation method, we will consider the extent to which the applicant presents a feasible, credible plan that includes the following:

(1) The type of design to be used (that is, random assignment or matched comparison). If matched comparison, include in the plan a discussion of why random assignment is not feasible.

(2) Outcomes to be measured.

(3) A discussion of how the applicant plans to assign students, teachers, classrooms, or schools to the project and control group or match them for comparison with other students, teachers, classrooms, or schools.

(4) A proposed evaluator, preferably independent, with the necessary background and technical expertise to carry out the proposed evaluation. An independent evaluator does not have any authority over the project and is not involved in its implementation.

In general, depending on the implemented program or project, under a competitive preference priority, random assignment evaluation methods will receive more points than matched comparison evaluation methods.

#### Definitions

As used in this notice—

*Scientifically based research* (section 9101(37) NCLB):

(A) Means research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs; and

(B) Includes research that—

(i) Employs systematic, empirical methods that draw on observation or experiment;

(ii) Involves rigorous data analyses that are adequate to test the stated hypotheses and justify the general conclusions drawn;

(iii) Relies on measurements or observational methods that provide

reliable and valid data across evaluators and observers, across multiple measurements and observations, and across studies by the same or different investigators;

(iv) Is evaluated using experimental or quasi-experimental designs in which individuals entities, programs, or activities are assigned to different conditions and with appropriate controls to evaluate the effects of the condition of interest, with a preference for random-assignment experiments, or other designs to the extent that those designs contain within-condition or across-condition controls;

(v) Ensures that experimental studies are presented in sufficient detail and clarity to allow for replication or, at a minimum, offer the opportunity to build systematically on their findings; and

(vi) Has been accepted by a peer-reviewed journal or approved by a panel of independent experts through a comparably rigorous, objective, and scientific review.

*Random assignment or experimental design* means random assignment of students, teachers, classrooms, or schools to participate in a project being evaluated (treatment group) or not participate in the project (control group). The effect of the project is the difference in outcomes between the treatment and control groups.

*Quasi experimental designs* include several designs that attempt to approximate a random assignment design.

*Carefully matched comparison groups design* means a quasi-experimental design in which project participants are matched with non-participants based on key characteristics that are thought to be related to the outcome.

*Regression discontinuity design* means a quasi-experimental design that closely approximates an experimental design. In a regression discontinuity design, participants are assigned to a treatment or control group based on a numerical rating or score of a variable unrelated to the treatment such as the rating of an application for funding. Eligible students, teachers, classrooms, or schools above a certain score (“cut score”) are assigned to the treatment group and those below the score are assigned to the control group. In the case of the scores of applicants’ proposals for funding, the “cut score” is established at the point where the program funds available are exhausted.

*Single subject design* means a design that relies on the comparison of treatment effects on a single subject or group of single subjects. There is little confidence that findings based on this

design would be the same for other members of the population.

*Treatment reversal design* means a single subject design in which a pre-treatment or baseline outcome measurement is compared with a post-treatment measure. Treatment would then be stopped for a period of time, a second baseline measure of the outcome would be taken, followed by a second application of the treatment or a different treatment. For example, this design might be used to evaluate a behavior modification program for disabled students with behavior disorders.

*Multiple baseline design* means a single subject design to address concerns about the effects of normal development, timing of the treatment, and amount of the treatment with treatment-reversal designs by using a varying time schedule for introduction of the treatment and/or treatments of different lengths or intensity.

*Interrupted time series design* means a quasi-experimental design in which the outcome of interest is measured multiple times before and after the treatment for program participants only.

#### Executive Order 12866

This notice of final priority has been reviewed in accordance with Executive Order 12866. Under the terms of the order, we have assessed the potential costs and benefits of this regulatory action.

The potential costs associated with the notice of final priority are those we have determined as necessary for administering applicable programs effectively and efficiently.

In assessing the potential costs and benefits—both quantitative and qualitative—of this notice of final priority, we have determined that the benefits of the final priority justify the costs.

We have also determined that this regulatory action does not unduly interfere with State, local, and tribal governments in the exercise of their governmental functions.

#### Intergovernmental Review

Some of the programs affected by this final priority are subject to Executive Order 12372 and the regulations in 34 CFR part 79. One of the objectives of the Executive order is to foster an intergovernmental partnership and a strengthened federalism. The Executive order relies on processes developed by State and local governments for coordination and review of proposed Federal financial assistance.

This document provides early notification of our specific plans and actions for these programs.

### Electronic Access to This Document

You may view this document, as well as all other Department of Education documents published in the **Federal Register**, in text or Adobe Portable Document Format (PDF) on the Internet at the following site: <http://www.ed.gov/news/fedregister>.

To use PDF you must have Adobe Acrobat Reader, which is available free at this site. If you have questions about using PDF, call the U.S. Government Printing Office (GPO), toll free, at 1-888-293-6498; or in the Washington, DC, area at (202) 512-1530.

**Note:** The official version of this document is the document published in the **Federal Register**. Free Internet access to the official edition of the **Federal Register** and the Code of Federal Regulations is available on GPO Access at: <http://www.gpoaccess.gov/nara/index.html>.

(Catalog of Federal Domestic Assistance Number does not apply.)

**Program Authority:** ESEA, as reauthorized by the No Child Left Behind Act of 2001, Pub. L. 107-110, January 8, 2002.

Dated: January 17, 2005.

**Rod Paige,**

*Secretary of Education.*

### Appendix—Analysis of Comments

*Comment:* Twenty-nine comments were received in support of the priority for random assignment studies of education policies and program interventions. Commenters noted that random assignment evaluations have been essential to understanding what works, what does not work, and what is harmful among interventions in many areas of public policy—including employment and training, welfare programs, health insurance, subsidies, pregnancy prevention, criminal justice, and substance abuse.

*Discussion:* The Secretary agrees with this comment.

*Change:* None.

*Comment:* One hundred and eighty-three respondents commented that random assignment is not the only method capable of generating understandings of causality. They stated that the Secretary's proposal would elevate experimental over quasi-experimental, observational, single-subject, and other designs which are sometimes more feasible and equally valid. However, 21 respondents commented that the priority correctly identifies random assignment experimental designs as the methodological standard for what constitutes scientific evidence for determining whether an intervention produces meaningful effects. The commenters pointed out that attempts to draw conclusions about intervention effects based on other methods have often led to misleading results. They stated that the priority is consistent with widely recognized

methodological standards in the social and medical sciences.

*Discussion:* The Secretary agrees that a random assignment design is not the only method capable of providing estimates of program effectiveness; however, it is the most defensible method in that it reliably produces an unbiased estimate of effectiveness. Conclusions about causality based on other methods, including the quasi-experimental designs included in this priority, have been shown to be misleading compared with experimental evidence. This is largely due to the difficulty in establishing equal treatment and comparison groups on all important characteristics related to the outcome variable with methods other than random assignment. The Secretary agrees with the latter commenters that random assignment is the standard for scientific evidence for determining the project effectiveness.

*Change:* None.

*Comment:* One hundred and seventy-three respondents commented that random assignment methods examine a limited number of isolated factors that are neither limited nor isolated in natural settings. These commenters stated that the complex nature of causality renders random assignment methods less capable of discovering causality than designs sensitive to local culture and conditions. Four respondents commented that random assignment methods estimate only the impact of the treatment and that the response to the treatment may vary according to contextual factors. These four respondents noted that random assignment assures that the contextual factors affecting outcomes are the same for the treatment and the control group and, therefore, the impact of the treatment is unambiguous. They noted further that it has not been demonstrated that evaluation methods "sensitive" to local culture and conditions can provide unambiguous answers as to whether the treatment is the cause of the observed outcome.

*Discussion:* The Secretary agrees with the latter comments. A major strength of the random assignment design is that it yields comparable treatment and control groups with respect to all characteristics and conditions, both observable and unobservable. When participants, e.g. students, teachers, classrooms, or schools, are randomly assigned to the project or to a control group, the only difference between the two groups is the impact of the treatment. While quasi-experimental designs, including carefully matched comparison groups, are also permitted under this priority, it is a practical impossibility to match on numerous characteristics and conditions, especially those that are unobservable or difficult to measure. However, case studies that collect information on local culture and conditions are an important complement to a random assignment study by providing a deeper understanding of the conditions that may influence the effectiveness of an intervention.

*Change:* None.

*Comment:* One hundred and eighty-six respondents commented that random assignment should sometimes be ruled out for reasons of ethics. For example, randomly assigning experimental subjects to

educationally inferior treatments, or denying control groups access to important instructional opportunities, is not ethically acceptable even when the results might be enlightening. Another 13 respondents commented that the priority recognizes that there are cases in which random assignment is not ethical and, in such cases, identifies quasi-experimental designs and single-subject designs as alternatives that may be justified by the circumstances of particular interventions.

*Discussion:* The Secretary agrees with both comments. There are occasions when random assignment is not an acceptable or feasible method of evaluation. The Department will address these issues in deciding whether or not to apply this priority in specific program competitions. Also, consistent with the American Psychological Association ethics code and in accordance with 34 CFR part 97, the Department has adopted the Common Rule for protection of human subjects in research including Subpart D dealing with inclusion of children in research. Grantees submit their plans for all research involving human subjects to an Institutional Review Board. All research involving human subjects must be conducted in accordance with an approved research protocol. This includes obtaining informed consent for participation when required by the Institutional Review Board as a condition of approval.

In general, random assignment does not pose ethical issues when employed to test the effectiveness of a new service or product that is believed to be beneficial and when the number of students who are equally eligible for and seeking that service is more than the number who can be served. When all applicants cannot be served, random assignment is fair, because it gives all participants an equal chance of being selected for the program.

When a random assignment evaluation is not ethical or not feasible, this priority includes quasi-experimental designs such as carefully matched comparison groups, regression discontinuity designs, single-subject designs, and interrupted time series that are capable of estimating program impacts. However, quasi-experimental designs do not provide the level of confidence in causal relationships that random assignment designs provide.

*Change:* None.

*Comment:* One hundred and seventy-four respondents commented that although it may be important to examine causality prior to wide implementation, pilot or exploratory programs are often too small in scale to provide reliable conclusions.

*Discussion:* The priority recognizes that for projects that are focused on special populations in which sufficient numbers of participants are not available to support random assignment or matched comparison group designs, single-subject designs such as multiple baseline or treatment-reversal or interrupted time series that are capable of demonstrating causal relationships can be employed. These small-scale or efficacy studies should lead to large-scale or effectiveness studies. Further, this priority is only relevant to programs for which demonstrations of effectiveness are

reasonable and relevant. The priority would generally not be applied in competitions to fund pilot or exploratory programs.

*Change:* None.

*Comment:* Two hundred and forty-two respondents commented that the choice of a research method must be determined by the goal or question being asked. They stated that alternative and mixed methods are rigorous and scientific and are important in knowing how well a program was implemented and what is "inside the box." Another group of 14 respondents commented that the priority does not preclude non-experimental designs, but gives clear priority to experimental designs for determining project effectiveness. These commenters noted that there may be areas in which an experimental design may not be feasible and non-experimental methods, including observational studies, may provide information on how to move research forward.

*Discussion:* The Secretary agrees with these comments. There are many research questions other than effectiveness that can be pursued. For these questions, research designs other than experimental and quasi-experimental would be appropriate. This priority is to be applied only when the question to be addressed is program effectiveness. The priority would be inappropriate if it were applied, for example, to applications in which the primary question is the fidelity of program implementation.

*Change:* None.

*Comment:* Twenty respondents expressed concern that the Department will make the priority a requirement for all grant competitions regardless of the intervention.

*Discussion:* The Secretary does not intend to make random assignment a requirement for all of the Department's grant competitions. The priority is intended for use only with discretionary grant programs in which grantees may use their funds to implement clearly specified interventions, and when the Department desires to obtain

evidence of the impact of those interventions on relevant outcomes.

*Change:* None.

*Comment:* One hundred and sixty-eight respondents disagreed with the Department's statement in the notice of proposed priority that "this regulatory action does not unduly interfere with State, local, and tribal governments in the exercise of their governmental functions." They took the position that as provision and support of programs are governmental functions so, too, is determining program effectiveness.

*Discussion:* As indicated above, the priority is for use only with discretionary grant programs in which awards are made on the basis of competition. The Secretary often establishes priorities for such programs and does not agree that supporting projects that would use scientific methods to evaluate the effectiveness of the interventions being implemented with grant funds would interfere with State, local, and tribal governments in the exercise of their governmental functions.

*Change:* None.

*Comment:* Six respondents expressed concern that the priority might limit what is studied or result in poorer quality programs being funded because of the additional points given to the evaluation priority.

*Discussion:* When using the priority to give competitive preference to an application, the Secretary intends to review applications using a two-stage process. The first stage would review the application without taking the priority into account. In the second stage of review, the applications rated highest in stage one would be reviewed for competitive preference. This will ensure that applications of lower program quality will not be funded as a result of additional points for the evaluation priority.

*Change:* Although no change has been made in the priority, the description of the competitive preference is clarified to include a two-stage review.

*Comment:* Nine respondents recommended that the Department continue to recognize the importance of independent evaluators.

*Discussion:* The priority gives preference to independent evaluators who have no authority over the project and are not involved in its implementation. Thus the importance of independent evaluators is recognized.

*Change:* None.

*Comment:* Twenty-three respondents expressed concern that there would be inadequate financial and technical resources in small programs and in rural areas to carry out a random assignment study and may prevent congressionally-intended beneficiary communities from receiving federal assistance.

*Discussion:* The priority provides for the use of alternate designs where insufficient numbers of participants are available to support random assignment or matched comparison group designs. The Secretary believes that investing in projects that generate evidence regarding the effectiveness of specified interventions would provide benefits beyond the individual grantee, and thus would represent a wise use of program dollars.

*Change:* None.

*Comment:* None.

*Discussion:* In order to make this priority more understandable to the general public, the Secretary believes that the priority would be improved by adding generally accepted definitions for technical terms used throughout the document. This may be helpful to practitioners and others who are interested in strengthening the evaluations of proposed projects but who may not be familiar with the specific types of evaluation described in this notice.

*Change:* The Secretary has added a definitions section to provide generally-accepted definitions of terms used throughout the document.

[FR Doc. 05-1317 Filed 1-24-05; 8:45 am]

BILLING CODE 4000-01-P