Peter Cade/Getty Images

THE PEW CHARITABLE TRUSTS | MacArthur Foundation

Results First Clearinghouse Database

User Guide

Contents

Overview

By using rigorous evidence to inform their decisions, policymakers can fund and operate public programs that have been proved to work and achieve substantially better results for citizens. Leveraging the power of evidence-based policymaking can help governments reduce wasteful spending, expand innovative programs, and strengthen accountability. The Pew-MacArthur Results First Initiative, a project of The Pew Charitable Trusts and the John D. and Catherine T. MacArthur Foundation, is working with states and counties to implement an innovative cost-benefit model to help direct funding to programs that can achieve strong outcomes.

To assist policymakers at all levels of government in identifying such programs and making data-driven budget decisions, the project has created the Results First Clearinghouse Database. This one-stop online resource provides policymakers with an easy way to find information on the effectiveness of various interventions as rated by eight national research clearinghouses. The database is not intended, however, to serve as a comprehensive list of all interventions that are worthy of funding. Rather, it is provided to assist policymakers as they consider the multitude of factors that inform budget and policy choices.

The clearinghouses conduct literature reviews and rate interventions in a range of policy areas based on rigorous evaluations. This vital information has the potential to play a key role in promoting evidence-based policymaking, but it is underutilized largely because each clearinghouse operates independently, uses somewhat different nomenclature when reporting results (e.g., rating programs as model, promising, potentially positive, etc.), and has limited capacity to publicize its work. Therefore, to use the information to make better investment choices, policymakers need to be aware that the clearinghouses exist, choose which to consult, find programs of interest, interpret the different ratings, and compare and contrast the findings.

To address the challenges posed by the existence of multiple clearinghouses, the Results First Clearinghouse Database has compiled information from eight clearinghouses in one place, reconciled the different systems and vocabularies, and provided the data in a clear, accessible format.

Clearinghouses included

The database contains information from eight national clearinghouses that conduct systematic research reviews to identify what works. While each uses slightly different procedures and criteria in its work, all use the same overall approach: They review and summarize research studies that assess the effectiveness of interventions.

Table 1 describes the areas of interest of each clearinghouse included in the database.

The user guide was updated in April 2017 to reflect new program information.

Table 1

Clearinghouses Included

Clearinghouse	Area of interest
Blueprints for Healthy Youth Development (http://www. blueprintsprograms.com)	**Youth** programs to prevent violence, crime, and drug use, as well as programs to improve mental and physical health, self-regulation, and educational achievement outcomes.
California Evidence-Based Clearinghouse for Child Welfare (http://www.cebc4cw.org)	**Child welfare** programs with an emphasis on those commonly used and marketed in California. Specific areas include: • Anger management, domestic violence, and substance abuse • Behavior management, including parent training • Core child welfare services, including placement and reunification • Engagement and parent partnering programs • Mental health • Prevention and early intervention • Support services for youth in the child welfare system
Coalition for Evidence-Based Policy (http://coalition4evidence. org)	**Social policy** interventions. Specific areas include: • Crime/violence prevention • Employment and welfare • Health care financing/delivery • Housing/homelessness • K-12 education • Mental health • Obesity prevention/treatment • Prenatal/early childhood • Substance abuse prevention/treatment • Teen pregnancy prevention
The U.S. Department of Justice's CrimeSolutions.gov (http://www. crimesolutions.gov)	**Criminal justice, juvenile justice, and crime victim services** programs and practices. Specific areas include: • Corrections and reentry • Courts • Crime and crime prevention • Drugs and substance abuse • Juveniles • Law enforcement • Technology and forensics • Victims and victimization

Table continues on next page

Clearinghouse	Area of interest
The U.S. Substance Abuse and Mental Health Services Administration's National Registry of Evidence-based Programs and Practices (http://www.nrepp. samhsa.gov)	Interventions that prevent and/or treat **mental and substance abuse disorders**. Specific areas include: • Mental health promotion • Mental health treatment • Substance abuse prevention • Substance abuse treatment • Co-occurring disorders
Promising Practices Network* (http://www.promisingpractices. net)	Programs that improve the lives of **children (prenatal to 18) and families**. Specific areas include: • Behavior problems • Child abuse and neglect • Cognitive development/school performance • Juvenile justice • Mental health • Physical health • Poverty/welfare • Substance use and dependence • Teen sex/pregnancy • Violent behavior
The Department of Education's What Works Clearinghouse (http://ies.ed.gov/ncee/wwc)	**Education** interventions. Specific areas include: • Children and youth with disabilities • College and career preparation • Dropout prevention • Early childhood education • English language learners • Literacy • Math • Postsecondary education • School choice • School organization and governance • Science • Student behavior • Teacher and leader effectiveness • Teacher incentives
What Works in Reentry (http://whatworks. csgjusticecenter.org)	**Reentry** interventions: programs, practices, or policies that affect at least one relevant outcome (recidivism, employment, or substance abuse) for people returning to the community from incarceration.

*   The Promising Practices Network was suspended in June 2014. Its website has been archived and will remain publicly available but will no longer be updated.

Information on interventions

The database includes information on more than 1,000 interventions and, for each, shows the policy area, intervention type (where applicable), clearinghouse that rated it, and Results First rating color (i.e., the colored dot). Users can click on the dot to view the clearinghouse's rating and a link to additional information. These elements are discussed in more detail below.

Policy area and intervention type

Each intervention is categorized in one of eight broad policy areas, although some have outcomes that affect multiple areas. For example, a substance abuse program for inmates is classified as adult criminal justice (as opposed to substance abuse), while a substance abuse program that targets teenagers in school is classified as substance abuse (as opposed to education).

For six of the policy areas, an additional classification, based on intervention type, is also used. For substance abuse and mental health, the intervention type denotes the target population, and for the other four areas it provides greater detail on program content.[1] Table 2 lists the policy areas and corresponding intervention types.

Table 2
Results First Clearinghouse Database Policy Areas and Intervention Types

Policy area	Intervention type
Adult criminal justice	None
Child welfare	Child welfare: general Child welfare: parenting
Education	Early childhood education K-12: dropout prevention K-12: general K-12: literacy K-12: math K-12: science K-12: special populations Postsecondary education
Health	Physical health Relationship strengthening Risky sexual behaviors/teen pregnancy Sexual/dating violence
Juvenile justice	None
Mental health	Mental health: adults Mental health: children Mental health: general

Table continues on next page

Policy area	Intervention type
Substance abuse	Substance abuse: adults Substance abuse: children Substance abuse: general
Youth development	Character education Classroom environment Prevention of risky behaviors Social competence Violence/bullying prevention

Clearinghouse ratings

Users can see the rating the clearinghouse assigned to an intervention (e.g., model, no effects, promising, etc.) by clicking on the colored dot. The clearinghouses use different methods for determining these ratings:

- **One overall rating for each intervention:** Blueprints for Healthy Youth Development, California Evidence-Based Clearinghouse for Child Welfare, Coalition for Evidence-Based Policy, and Promising Practices Network use this approach. The database displays the overall rating.

- **One overall rating for each outcome of an intervention:** National Registry of Evidence-based Programs and Practices and What Works Clearinghouse use this method.[2] The database displays the highest rating given to any of the intervention's outcomes. As such, the rating for only one of the intervention's outcomes is shown.

- **Individual ratings for specific outcomes:** What Works in Reentry Clearinghouse, which employs this system, assigns ratings to three outcomes: recidivism, employment, and substance abuse. The database displays the highest rating given to any of these.[3] In addition, What Works in Reentry Clearinghouse assesses the rigor of each evaluation as either high or basic.

- **One overall rating for interventions and individual ratings for each outcome:** CrimeSolutions.gov is the only clearinghouse that uses this dual approach. It does so because it reviews both programs and practices, which its website defines as "a general category of programs, strategies, or procedures that share similar characteristics with regard to the issues they address and how they address them." For programs, CrimeSolutions.gov assigns an overall rating; for practices, it assigns one rating to each outcome.[4] Subsequently, the database displays the overall rating for programs and the highest rating given to any of the outcomes for practices.

It is important to keep in mind that while the clearinghouses use the same overall approach, there are differences in their rating terminology, criteria, and processes. (See Table 3 for the terminology used by each clearinghouse.) In general, the clearinghouses' ratings reflect both the rigor of the evidence and the significance of the impact.[5] They also are all based on studies that use randomized control trials or strong quasi-experimental designs and have outcome measures that are valid and reliable. (For more on the rating criteria, see the appendix.)

Results First rating colors

For each intervention, the database displays a colored dot. This color, which is based on the clearinghouse's own rating system, indicates where each intervention falls on a spectrum from highest rated (green) to negative effects (red). The rating color also enables users to make a cursory comparison among clearinghouses' ratings for an intervention. However, the rating color does not indicate that the clearinghouses' methodologies are identical.

For example, green denotes similar standards of evidence for Blueprints for Healthy Youth Development and California Evidence-Based Clearinghouse for Child Welfare, but the criteria used by each clearinghouse are somewhat distinct.

Table 3 shows the clearinghouse ratings and the colors Results First assigned them.

Table 3
Clearinghouse Ratings

Clearinghouse	Highest rated	Second-highest rated	No evidence of effects	Mixed effects	Negative effects
Blueprints for Healthy Youth Development	Model plus; and model	Promising	N/A	N/A	N/A
California Evidence-Based Clearinghouse for Child Welfare	1: Well-supported by research evidence 2: Supported by research evidence*	3: Promising research evidence	4: Evidence fails to demonstrate effect	N/A	5: Concerning practice
Coalition for Evidence-Based Policy	Top tier	Near top tier	N/A	N/A	N/A
CrimeSolutions.gov	Effective	Promising	No effects†	N/A	N/A
National Registry of Evidence-based Programs and Practices‡	Effective	Promising	Ineffective§	N/A	N/A
	Score of 3.0-4.0	Score of 2.0-2.9	N/A	N/A	N/A
Promising Practices Network	Proven; and proven/promising‖	Promising	N/A	N/A	N/A
What Works Clearinghouse	Positive	Potentially positive	No discernible effects	Mixed effects	Potentially negative; and negative
What Works in Reentry Clearinghouse#	Strong beneficial	Modest beneficial	No evidence of effect	N/A	Modest harmful; and strong harmful

* The California Evidence-Based Clearinghouse for Child Welfare has three ratings for positive findings. To fit them within the color categorization, two ratings were combined. Results First decided to classify the top two ratings as green (highest rated) because their definitions most closely matched those of the other clearinghouses' ratings for this color.

† "No effects" includes interventions found to have either no or harmful effects.

‡ In November 2015, the National Registry of Evidence-based Programs and Practices began to review programs under new guidelines. The new ratings are shown in the first row. The legacy ratings, which are still available on its website, are shown in the second row.

§ "Ineffective" includes interventions found to have either no or potentially harmful effects.

‖ "Proven/promising" refers to interventions that received a proven rating for one indicator and a promising rating for another.

The terminology used in the database for What Works in Reentry Clearinghouse is an abbreviated version of its nomenclature due to space limitations.

As shown in Table 3, only What Works Clearinghouse uses the mixed effects category. Further, Blueprints for Healthy Youth Development, Coalition for Evidence-Based Policy, and Promising Practices Network assign ratings solely to interventions that have positive effects (i.e., highest rated and second-highest rated).

Table 4 shows the rating colors, corresponding categories, and general definitions.

Table 4
Results First Rating Colors

Rating color	Rating category	Broad definition
🟢	**Highest rated**	The clearinghouse assigned the intervention its highest rating. In general, this requires one to two evaluations that: a) use the strongest research designs, including randomized control trials or high-quality quasi-experimental designs; and b) show that the intervention had a statistically significant positive impact.
🟡	**Second-highest rated**	The clearinghouse assigned the intervention its second-highest rating. This generally requires an evaluation that used a quasi-experimental design and showed that the intervention had a positive impact.
⚪	**No evidence of effects**	The clearinghouse found the intervention to have no statistically significant effects based on at least one evaluation that used a randomized control trial or a quasi-experimental design.
🔵	**Mixed effects**	This category is only used for interventions reviewed by What Works Clearinghouse. It is applied when two or more evaluations using randomized control trials or quasi-experimental designs found inconsistent effects of a given intervention. That is, one study showed a positive effect and either another showed a negative effect or at least two others showed indeterminate effects.
🔴	**Negative effects**	The clearinghouse determined that the intervention had negative effects, which could be statistically significant or not, based on at least one evaluation that used a randomized control trial or a quasi-experimental design.

Note: "Statistically significant" refers to whether the effect being studied was due to chance, i.e., where the p-value is less than or equal to .05.

Hyperlink to program pages

By clicking on "Learn more" in the callout box, users can access a hyperlink to the relevant clearinghouse program page. These pages provide key information, such as a description of the intervention, study findings, outcomes examined, evaluation methodology, and references. They may also include other details, such as costs, target population, additional ratings, and implementation resources.

How to use the database

The database is intended to help policymakers identify whether the programs they fund or are considering funding are evidence-based and, if so, their level of effectiveness.

Step 1: Find an intervention

Users can search the database in a number of ways: by title, policy area, intervention type (if applicable), clearinghouse, rating color, or a combination of these. Users can also leave the search options blank to access all interventions in the database.

Results First recommends that users keep their searches as broad as possible because interventions can target multiple policy areas but are categorized in only one. Programs also may be reviewed by more than one clearinghouse.

Step 2: Review the intervention's description

Once an intervention that matches the name of the program has been found, determine whether it is a true match. Do this by first clicking on the colored dot and then on "Learn more." This takes the user to the clearinghouse's program page, which contains a detailed description of the intervention. Results First recommends that users review all relevant pages because each provides a unique description.

Step 3: Review the ratings

Look at the ratings assigned to the intervention by the clearinghouses, which are provided in the callout boxes and on the clearinghouses' program pages. Roughly 65 percent of the interventions have only one rating; the other 35 percent have more, meaning they were reviewed by more than one clearinghouse. Among those with multiple ratings:

- 40 percent received the same rating category from all clearinghouses that reviewed them. This is demonstrated by the fact that the color for each of these interventions is identical across all clearinghouses (e.g., all green or all yellow).

- 50 percent have ratings that are a mix of highest rated and second-highest rated, as evidenced by their green and yellow colors. This indicates that the clearinghouses agreed the interventions were evidence-based and produced positive effects but disagreed on the strength of the evidence.

- 10 percent were given diverse ratings by the clearinghouses, which are represented with multiple colors. In some cases, this was because the clearinghouses based their assessments on different outcomes of the intervention. In other cases, it was a result of their distinct methodologies, including the research they reviewed.

Step 4: Use the ratings to inform decision-making

How ratings will be applied to policy and budget decisions will depend on the needs of the users and the context of their situations. For instance, this information can be taken into account when formulating agency budgets, allocating resources to different programs within a budget, and considering whether to fund a program. In general, interventions with higher ratings (e.g., highest rated vs. second-highest rated) are more likely to produce positive outcomes when delivered at scale.

The database can also be used to determine the percentage of currently funded programs in a policy area that is evidence-based. More specifically, it can establish a baseline that enables government leaders to see which funded programs have strong evidence that they are working (i.e., highest rated), which have some evidence of positive impact (i.e., second-highest rated), and which have limited or no evidence on their outcomes (i.e., no effects or negative effects). This can be accomplished by developing an inventory of the programs that a department or agency funds and then using the database to assess the available evidence of effectiveness for each one.

Although the clearinghouse database is meant to help policymakers make data-driven budget choices, it should not be viewed as a comprehensive list of all interventions that are worthy of funding. Policymakers consider a multitude of other factors when making these decisions, such as the populations to be served, the risk and need levels of those populations, local information on the interventions' effectiveness, the availability of implementation materials, staff capacity, and resources.

FAQs

What are national clearinghouses?

In general, the clearinghouses are charged with determining what works based on rigorous reviews of evidence. While all clearinghouses use their own criteria and procedures, they share an overall approach. Each reviews research studies in its area of interest (education, criminal justice, etc.) and determines whether there is sufficient evidence to support a conclusion about an intervention's effectiveness. The clearinghouses assign ratings (e.g., model, promising, mixed effects) to interventions that meet the evidence standard. Thus, only evidence-based interventions are listed by the clearinghouses and included in the clearinghouse database. However, it should be noted that some interventions have no effect or a negative impact. For example, CrimeSolutions.gov found Ending Violence to have no effect while What Works in Reentry Clearinghouse found Project Greenlight to have a harmful effect. Three other clearinghouses—Blueprints for Healthy Youth Development, Coalition for Evidence-Based Policy, and Promising Practices Network—report only interventions that have positive effects.

How do I find out which outcomes of interventions were evaluated?

Click on "Learn more" within each callout box to access the clearinghouses' information on their program pages. Most clearinghouses clearly lay out the outcomes they assessed for each intervention and the impact the intervention had on them. The "policy area" column also can be helpful: It shows which broad area the key outcomes affect, such as youth development. The intervention type provides even greater detail, such as violence and bullying prevention.

The database does not display individual outcomes, because doing so would make the database unwieldy and difficult to read. Because clearinghouses report one to 10 outcomes per intervention, it was not feasible to present all of that information in a clear and useful way.

Why aren't all interventions rated by multiple clearinghouses?

Each clearinghouse has a specific focus, such as criminal justice, child welfare, or education. As a result, some interventions fall outside of a clearinghouse's purview. Additionally, each clearinghouse has a somewhat different process for choosing interventions to review. For instance, Blueprints for Healthy Youth Development conducts an ongoing literature search for studies of youth programs, while California Evidence-Based Clearinghouse for Child Welfare examines only interventions that fit under topic areas chosen for that fiscal year. Lastly, each clearinghouse has different requirements concerning its research base. California Evidence-Based Clearinghouse for Child Welfare looks only at published, peer-reviewed research. All other clearinghouses consider evaluations included in peer-reviewed publications or comprehensive evaluation reports.

How can clearinghouses assign different ratings to the same intervention?

There are several reasons for differences among ratings. First, each clearinghouse has its own rating system and assigns slightly different weights to various aspects of evaluations. Second, some clearinghouses include components in their rating systems that do not focus specifically on methodology or impact. For instance, CrimeSolutions.gov and National Registry of Evidence-based Programs and Practices include program fidelity in their scoring. Lastly, as described above, clearinghouses often examine different research for the same intervention.

How did Results First determine which clearinghouses to include in its database?

In general, the database includes clearinghouses that operate at a national level (either as part of a federal agency or a national organization), are well-known, examine policy areas that mirror those in the Results First benefit-cost model, and employ a rating system that is based on a strong methodology.

How frequently will the Results First Clearinghouse Database be updated?

The database will be updated semiannually to reflect new programs rated by the clearinghouses.

Appendix

Clearinghouse rating criteria

The table below describes and defines the ratings used by each clearinghouse. Additional information regarding the ratings systems can be found on their websites.

Clearinghouse	Rating system
Blueprints for Healthy Youth Development*	**Model plus, model, and promising:** Each must meet the following criteria: • **Intervention specificity:** The program description clearly identifies the outcome the program is designed to change, the specific risk and/or protective factors targeted to produce this change in outcome, the population for which it is intended, and how the components of the intervention work to produce this change. • **Evaluation quality:** The evaluation trials produce valid and reliable findings. **Model plus and model programs** require a minimum of (a) two high-quality, randomized control trials (RCTs) or (b) one high-quality, RCT plus one high-quality, quasi-experimental evaluation. A **promising program** requires a minimum of (a) one high-quality, RCT or (b) two high-quality, quasi-experimental evaluations. • **Intervention impact:** The preponderance of evidence from the high-quality evaluations indicates significant positive change in intended outcomes that can be attributed to the program and there is no evidence of harmful effects. For **model plus and model programs**, positive intervention impact must be sustained for a minimum of 12 months after the program intervention ends. • **Dissemination readiness:** The program is available for dissemination and has the necessary organizational capability, manuals, training, technical assistance, and other support required for implementation with fidelity in communities and public service systems. Model plus programs also must meet the following criterion: • **Independent replication:** In at least one high-quality study demonstrating desired outcomes, authorship, data collection, and analysis have been conducted by a researcher who is neither a current or past member of the program developer's research team and who has no financial interest in the program.
California Evidence-Based Clearinghouse for Child Welfare[†]	The Scientific Rating Scale is a 1-to-5 rating of the strength of the research evidence supporting a practice or program. A scientific rating of 1 represents a practice with the strongest research evidence, and a 5 represents a concerning practice that appears to pose substantial risk to children and families. **1 = Well-supported by research evidence** • At least two rigorous RCTs, in different usual care or practice settings have found the practice to be superior to an appropriate comparison practice. • In at least one RCT, the practice has been shown to have a sustained effect at least one year beyond the end of treatment. **2 = Supported by research evidence** • At least one rigorous RCT in usual care or a practice setting has found the practice to be superior to an appropriate comparison practice. • In at least one RCT, the practice has shown to have a sustained effect at least six months beyond the end of treatment. **3 = Promising research evidence** • At least one study utilizing some form of control has established the practice's benefit over the control, or found it to be comparable to a practice rated 1, 2, or 3 on this rating scale or superior to an appropriate comparison practice. • If multiple outcome studies have been conducted, the overall weight of evidence supports the benefit of the practice.

Table continues on next page

Clearinghouse	Rating system
California Evidence-Based Clearinghouse for Child Welfare[†] **(continued)**	**In addition for 1, 2, and 3** • There are no case data suggesting a risk of harm that was a) probably caused by the treatment and b) severe or frequent. • There is no legal or empirical basis suggesting that, compared with its likely benefits, the practice constitutes a risk of harm to those receiving it. • The practice has a book, manual, and/or other written guidelines that specify the components of the practice protocol and describe how to administer it. • Studies must have been reported in published, peer-reviewed literature. • The overall weight of evidence supports the benefit of the practice. **4 = Evidence fails to demonstrate effect** • Two or more RCTs have found that the practice has not resulted in improved outcomes when compared with usual care. • The overall weight of evidence does not support the benefit of the practice. **5 = Concerning practice** • The overall weight of evidence suggests the intervention has a negative effect upon clients served; and/or • There are no case data suggesting a risk of harm that was a) probably caused by the treatment and b) severe or frequent; and/or • There is a legal or empirical basis suggesting that, compared with its likely benefits, the practice constitutes a risk of harm to those receiving it.
Coalition for Evidence-Based Policy[‡]	**Top tier:** Interventions, preferably conducted in typical community settings, have been shown in well-conducted RCTs to produce sizable, sustained benefits to participants and/or society. **Near top tier:** Interventions have been shown to meet almost all elements of the top tier standard and need only one additional step to qualify. This category includes, for example, interventions that meet all elements of the standard in a single site and need only a replication trial to confirm the initial findings and establish that they can be generalized to other sites. Checklists, which closely track guidance from both the U.S. Office of Management and Budget and the U.S. Education Department's Institute of Education Sciences, are used to determine whether the RCTs are well-conducted. These lists assess the overall study design, whether the intervention and control groups remained equivalent during the study, the study's outcome measures, and the study's reporting of the intervention's effects.
CrimeSolutions.gov[§]	CrimeSolutions.gov uses three ratings: **effective, promising, and no effects**. Programs and practices each have different requirements for them. **Programs** Each must meet the following criteria: • The program must be evaluated with at least one RCT or quasi-experimental research design (with a comparison condition). • The outcomes assessed must relate to crime, delinquency, or victimization prevention, intervention, or response. • The evaluation must be published in a peer-reviewed publication or documented in a comprehensive evaluation report. • The date of publication must be 1980 or later.

Table continues on next page

Clearinghouse	Rating system
CrimeSolutions.gov[§]	• **Effective:** Programs have strong evidence to indicate they achieve their intended outcomes when implemented with fidelity. Requires at least one very rigorous and well-designed study that finds significant, positive effects on justice-related outcomes; and no studies that find significant, harmful effects on justice-related outcomes. • **Promising:** Programs have some evidence to indicate they achieve their intended outcomes. Requires at least one well-designed but slightly less rigorous study that finds significant, positive effects on justice-related outcomes; and no studies that find significant, harmful effects on justice-related outcomes. • **No effects:** Programs have strong evidence indicating that they had no effects or had harmful effects when implemented with fidelity. Requires at least one very rigorous and well-designed study that finds either significant harmful effects or no significant effects on justice-related outcomes. **Practices** CrimeSolutions.gov assigns each outcome a rating, but the database reports only the highest-rated outcome. Practices rely on meta-analyses instead of evaluations of individual programs. Each meta-analysis must meet the following criteria: • It includes and aggregates the results of at least two studies. • It reports on at least one eligible outcome related to crime, delinquency, overt problem behaviors (e.g., aggression, gang involvement, substance abuse), crime victimization, justice system practices or policies, or risk factors for crime and delinquency. • All studies included in the meta-analysis must include an appropriate control, comparison or counterfactual condition, or the meta-analysis must analyze these studies separately from those that appropriate counterfactuals. • It reports effect sizes that represent the magnitude of the treatment effect. • At least 50 percent of the studies included in the meta-analysis must be published or otherwise available on or after 1980. • Samples included in the meta-analysis must be restricted to either adults or juveniles, or mean effect sizes for adults and juveniles must be reported separately. Each meta-analysis is then scored for overall quality, and each outcome is assessed for internal validity. The results, along with information about the direction and statistical significance of the mean effect size, are combined to produce the following outcome ratings: • **Effective:** The highest-quality evidence shows the outcome had a statistically significant positive effect. • **Promising:** Moderate-quality evidence shows the outcome had a statistically significant positive effect. • **No effects:** Moderate- to high-quality evidence shows the outcome had no statistically significant effect or a statistically significant negative effect.
National Registry of Evidence-based Programs and Practices[**]	In November 2015, the National Registry of Evidence-based Programs and Practices instituted new guidelines for reviewing programs. The programs reviewed under the previous criteria (now referred to as "legacy" programs) will continue to appear on its website until they have been re-reviewed under the new guidelines, which is expected to take four years (2015 through 2018). The following describes both rating systems.

Table continues on next page

Clearinghouse	Rating system
National Registry of Evidence-based Programs and Practices** (continued)	**Current Rating System** The National Registry of Evidence-based Programs and Practices assigns each outcome one of the following ratings: **effective, promising, ineffective, or inconclusive**. The database reports only the highest-rated outcome but does not include programs with outcomes rated only as inconclusive. Each intervention must first meet the following requirements: • Research or evaluation of the intervention has assessed mental health or substance use outcomes among individuals, communities, or populations OR other behavioral health-related outcomes on individuals, communities, or populations with or at risk of mental health issues or substance use problems. • Evidence of these outcomes has been demonstrated in at least one study using an experimental or quasi-experimental design. • Within the previous 25 years, the results of these studies have been published in a peer-reviewed journal or other professional publication, or documented in a comprehensive evaluation report. Then, outcomes are rated on four dimensions: rigor, program fidelity, effect size, and conceptual framework. The results are combined to produce the following outcome ratings: • **Effective:** The evidence base produced strong evidence of a favorable effect. • **Promising:** The evidence base produced sufficient evidence of a favorable effect. • **Ineffective:** The evidence base produced sufficient evidence of a negligible effect or a possibly harmful effect. • **Inconclusive:** Limitations in the study design or a lack of effect size information preclude from reporting further on the effect. **Legacy Rating System** Each intervention must first meet the following requirements: • The intervention has produced one or more positive behavioral outcomes ($p \leq .05$) in mental health or substance abuse among individuals, communities, or populations. Significant differences between groups over time must be demonstrated for each outcome. • Evidence of positive behavioral outcome(s) has been demonstrated in at least one study using an experimental or quasi-experimental design. • The results of these studies have been published in a peer-reviewed journal or other professional publication or documented in a comprehensive evaluation report. • Implementation materials, training and support resources, and quality assurance procedures have been developed and are ready for use by the public. Then, the outcomes are separately scored from 0 to 4.0 on the following six criteria related to the quality of research: **Reliability of measures** • 0 = Absence of evidence of reliability or evidence that some relevant types of reliability did not reach acceptable levels. • 2 = All relevant types of reliability have been documented to be at acceptable levels in studies by the applicant. • 4 = All relevant types of reliability have been documented to be at acceptable levels in studies by independent investigators.

Table continues on next page

Clearinghouse	Rating system
National Registry of Evidence-based Programs and Practices** **(continued)**	**Validity of measures** • 0 = Absence of evidence of measure validity, or some evidence that the measure is not valid. • 2 = Measure has face validity; absence of evidence that measure is not valid. • 4 = Measure has one or more acceptable forms of criterion-related validity (correlation with appropriate, validated measures or objective criteria); or, for objective measures of response, there are procedural checks to confirm data validity; absence of evidence that measure is not valid. **Intervention fidelity** • 0 = Absence of evidence or only narrative evidence that the applicant or provider believes the intervention was implemented with acceptable fidelity. • 2 = There is evidence of acceptable fidelity in the form of judgment(s) by experts, systematic collection of data (e.g., dosage, time spent in training, adherence to guidelines or a manual), or a fidelity measure with unspecified or unknown psychometric properties. • 4 = There is evidence of acceptable fidelity from a tested fidelity instrument shown to have reliability and validity. **Missing data and attrition** • 0 = Missing data and attrition were taken into account inadequately, or there was too much to control for bias. • 2 = Missing data and attrition were taken into account by simple estimates of data and observations, or by demonstrations of similarity between remaining participants and those lost to attrition. • 4 = Missing data and attrition were taken into account by more sophisticated methods that model missing data, observations, or participants, or there was no attrition or missing data needing adjustment. **Potential confounding variables** • 0 = Confounding variables or factors were as likely to account for the outcome(s) reported as were the hypothesized causes. • 2 = One or more potential confounding variables or factors were not completely addressed, but the intervention appears more likely than these confounding factors to account for the outcome(s) reported. • 4 = All known potential confounding variables appear to have been completely addressed in order to allow causal inference between the intervention and outcome(s) reported. **Appropriateness of analysis** • 0 = Analyses were not appropriate for inferring relationships between intervention and outcome, or sample size was inadequate. • 2 = Some analyses may not have been appropriate for inferring relationships between intervention and outcome, or sample size may have been inadequate. • 4 = Analyses were appropriate for inferring relationships between intervention and outcome. **Sample size and power were adequate.** Each outcome also receives an overall (average) quality of research score, which is shown in the clearinghouse database. It is also what Results First used to determine the rating color: 2.0 to 3.0 is yellow (second-highest rated), and 3.0 to 4.0 is green (highest rated). Programs with ratings below 2.0 were not included because they would not have met the methodological standards of any other clearinghouse.

Table continues on next page

Clearinghouse	Rating system
Promising Practices Network[††]	**Proven, proven/promising, and promising.** These ratings are based on the scores each study receives in the areas below: **1. Type of outcomes affected** • Proven: Program directly affects one of the Promising Practices Network's indicators. • Promising: Program affects an intermediary outcome for which there is evidence that it is likely to be related to one of the Promising Practices Network's indicators. **2. Substantial effect size** • Proven: At least one outcome is changed by 20 percent, 0.25 standard deviations, or more. • Promising: Change in outcome is more than 1 percent. **3. Statistical significance** • Proven: At least one outcome with a substantial effect size is statistically significant at the 5 percent level. • Promising: Outcome change is significant at the 10 percent level. **4. Research design** • Proven: Study uses a convincing comparison group to identify program impact, including an RCT or quasi-experimental design with matched control, and uses statistical methods to control for self-selection bias that might otherwise influence outcomes. • Promising: Study has a comparison group, but it suffers from some weaknesses (e.g., the groups lack comparability on preexisting variables or the analysis does not employ appropriate statistical controls). **5. Sample size** • Proven: Sample size exceeds 30 in both treatment and comparison groups, or sample sizes are large enough to provide power to detect effects. • Promising: Sample size of evaluation exceeds 10 in both the treatment and comparison groups. Promising Practices Network was suspended indefinitely in June 2014. The website has been archived and will remain publicly available but will no longer be updated. Note: Rating if one indicator is "proven" and another indicator is "promising."
What Works Clearinghouse[‡‡]	What Works Clearinghouse assigns each outcome one of the following ratings: **positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.** The database reports only the highest-rated outcome. All studies reviewed must meet What Works Clearinghouse standards without reservations (an RCT with low attrition) or with reservations (RCT with high attrition and/or quasi-experimental design with baseline equivalence). The below terminology is used to define the ratings: • Statistically significant positive: The estimated effect is positive and statistically significant (correcting for clustering when not properly aligned). • Substantively important positive effect: The estimated effect is positive and not statistically significant but is substantively important. • Indeterminate effect: The estimated effect is neither statistically significant nor substantively important. • Substantively important negative effect: The estimated effect is negative and not statistically significant but is substantively important.

Table continues on next page

Clearinghouse	Rating system
What Works Clearinghouse ‡‡ **(continued)**	• Statistically significant negative effect: The estimated effect is negative and statistically significant (correcting for clustering when not properly aligned). Note: A statistically significant estimate of an effect is one for which the probability of observing such a result by chance is less than 1 in 20 (using a two-tailed t-test with p=.05). A properly aligned analysis is one for which the unit of assignment and unit of analysis are the same. An effect size of 0.25 standard deviations or larger is considered to be substantively important. Interventions must meet all of the following conditions to receive the relevant rating: **Positive** • Two or more studies show statistically significant positive effects, at least one of which meets What Works Clearinghouse group design standards without reservations. • No studies show statistically significant or substantively important negative effects. **Potentially positive** • At least one study shows statistically significant or substantively important positive effects. • Fewer or the same number of studies show indeterminate effects than show statistically significant or substantively important positive effects. • No studies show statistically significant or substantively important negative effects. **No discernible effects** • None of the studies shows statistically significant or substantively important effects, either positive or negative. **Mixed** • At least one study shows statistically significant or substantively important positive effects. • At least one study shows statistically significant or substantively important negative effects, but no more such studies than the number showing statistically significant or substantively important positive effects. Or • At least one study shows statistically significant or substantively important effects. • More studies show an indeterminate effect than show statistically significant or substantively important effects. **Potentially negative** • One study shows statistically significant or substantively important negative effects. • No studies show statistically significant or substantively important positive effects. Or • Two or more studies show statistically significant or substantively important negative effects, and at least one study shows statistically significant or substantively important positive effects. • More studies show statistically significant or substantively important negative effects than show statistically significant or substantively important positive effects. **Negative** • Two or more studies show statistically significant negative effects, at least one of which meets What Works Clearinghouse group design standards without reservations. • No studies show statistically significant or substantively important positive effects.

Table continues on next page

Clearinghouse	Rating system
What Works in Reentry Clearinghouse[§§]	There are two levels of ratings. The rigor of each study is rated as high or basic. Then each outcome (recidivism, employment, and/or substance abuse) for each study is rated on the evidence of effectiveness. **Rigor** All studies (high and basic) must meet the following requirements: • Be an RCT or a quasi-experimental design with matched groups and/or statistical controls for group differences. • Have minimal threats to validity. • Have been conducted by an independent researcher or published in a peer-reviewed journal. • Be published in or after 1980. **High** • At least 100 participants in both treatment and control groups; total number of participants at least 200. • Attrition must not present a significant threat to the study's validity. • Outcomes must be tracked for at least one year following the individual's completion of the program or release from incarceration. **Basic** • At least 30 in both treatment and control groups; total number of participants at least 60. **<u>Evidence of effectiveness</u>** • **Strong beneficial evidence (strong beneficial):** The study findings show a consistent pattern indicating that the treatment group experienced better post-release outcomes in the area of interest than the comparison group, and most or all of these findings are statistically significant. • **Modest beneficial evidence (modest beneficial):** Study findings are mixed: Some indicate that the treatment group experienced significantly better outcomes than the comparison group, but other findings show no significant differences; or findings indicate that the treatment group experienced better outcomes on several measures, but these differences generally failed to reach statistical significance. • **No evidence of an effect (no evidence of effect):** The study findings show very few or no significant differences between groups. • **Modest harmful evidence (modest harmful):** Study findings are mixed: Some indicate that the comparison group experienced significantly better outcomes than the treatment group, but other findings show no significant differences; or findings indicate that the comparison group experienced better outcomes on several measures, but these differences generally failed to reach statistical significance. • **Strong harmful evidence (strong harmful):** The study findings show a consistent pattern indicating that the comparison group experienced better post-release outcomes than the treatment group, and most or all of these findings are statistically significant.

* Blueprints for Healthy Youth Development, http://www.blueprintsprograms.com/programCriteria.php.

† The California Evidence-Based Clearinghouse for Child Welfare, http://www.cebc4cw.org/ratings/scientific-rating-scale.

‡ Coalition for Evidence-Based Policy, http://toptierevidence.org/solicitationreview-process.

§ CrimeSolutions.gov, http://www.crimesolutions.gov/about_starttofinish.aspx.

** National Registry of Evidence-based Programs and Practices, http://www.nrepp.samhsa.gov/04a_review_process.aspx.

†† Promising Practices Network, http://www.promisingpractices.net/criteria.asp.

‡‡ What Works Clearinghouse, http://ies.ed.gov/ncee/wwc/InsidetheWWC.aspx#process.

§§ What Works in Reentry Clearinghouse, http://csgjusticecenter.org/reentry/about-what-works.

Endnotes

1   Results First plans to add intervention types to adult criminal justice and juvenile justice in the future.

2   What Works Clearinghouse reviews interventions according to topic area (e.g., dropout prevention, elementary school math, and high school math), and as a result, one intervention can be reviewed multiple times. For instance, Accelerated Reader was reviewed in the areas of beginning reading and adolescent literacy. To depict this in the database, there are separate entries for Accelerated Reader (adolescent literacy) and Accelerated Reader (beginning reading). What Works Clearinghouse rates outcomes within each of these areas. The database displays the highest-rated outcome for each topic area.

3   "Treatment Alternatives to Street Crime (TASC)," What Works in Reentry Clearinghouse, accessed July 3, 2014, http://whatworks. csgjusticecenter.org/interventions/treatment-alternatives-to-street-crime-tasc. In most cases, each outcome has only one rating. However, if What Works in Reentry Clearinghouse reviewed multiple evaluations that used the same study sample, then the outcome would be assigned a rating for every evaluation. A good example is TASC. What Works in Reentry Clearinghouse reviewed two studies for this program and gave the recidivism outcome two ratings: no evidence of effect: basic rigor; and modest beneficial: basic rigor.

4   To differentiate between the two types of reviews, the database displays "(practice)" after the intervention name where applicable.

5   For instance, What Works Clearinghouse rates an outcome as "positive" if the intervention's evaluation was a randomized control trial with low attrition (i.e., rigor of the evidence) and showed that the outcome had a statistically significant positive effect (i.e., significance of the impact).

The Pew Charitable Trusts

901 E St. NW, 10th Floor
Washington, DC 20004

pewstates.org

The Pew Charitable Trusts is driven by the power of knowledge to solve today's most challenging problems. Pew applies a rigorous, analytical approach to improve public policy, inform the public, and invigorate civic life.

The John D. and Catherine T. MacArthur Foundation

140 S. Dearborn St.
Chicago, IL 60603

macfound.org

The John D. and Catherine T. MacArthur Foundation supports creative people, effective institutions, and influential networks building a more just, verdant, and peaceful world. MacArthur is placing a few big bets that truly significant progress is possible on some of the world's most pressing social challenges, including over-incarceration, global climate change, nuclear risk, and significantly increasing financial capital for the social sector. In addition to the MacArthur Fellows Program, the Foundation continues its historic commitments to the role of journalism in a responsible and responsive democracy, as well as the strength and vitality of our headquarters city, Chicago.