

JOINT STATE GOVERNMENT COMMISSION

General Assembly of the Commonwealth of Pennsylvania

EVIDENCE-BASED DECISION MAKING BY THE COMMONWEALTH OF PENNSYLVANIA

A STAFF STUDY

JULY 2018



*Serving the General Assembly of the
Commonwealth of Pennsylvania Since 1937*

REPORT

Evidence-Based Decision Making By The Commonwealth Of Pennsylvania

Project Managers:	Glenn Pasewicz, Executive Director
Project Staff:	Michael Dirckx, Staff Attorney Bryan DeWalt, Public Policy Analyst Amr Hassanein, Public Policy Intern Kahla Lukens, Administrative Assistant

JOINT STATE GOVERNMENT COMMISSION

Room 108 Finance Building
613 North Street
Harrisburg, PA 17120-0108

Telephone: 717-787-4398
Fax: 717-783-9380
E-mail: jntst02@legis.state.pa.us
Website: <http://jsg.legis.state.pa.us>

The Joint State Government Commission was created in 1937 as the primary and central non-partisan, bicameral research and policy development agency for the General Assembly of Pennsylvania.¹

A fourteen-member Executive Committee comprised of the leadership of both the House of Representatives and the Senate oversees the Commission. The seven Executive Committee members from the House of Representatives are the Speaker, the Majority and Minority Leaders, the Majority and Minority Whips, and the Majority and Minority Caucus Chairs. The seven Executive Committee members from the Senate are the President Pro Tempore, the Majority and Minority Leaders, the Majority and Minority Whips, and the Majority and Minority Caucus Chairs. By statute, the Executive Committee selects a chairman of the Commission from among the members of the General Assembly. Historically, the Executive Committee has also selected a Vice-Chair or Treasurer, or both, for the Commission.

The studies conducted by the Commission are authorized by statute or by a simple or joint resolution. In general, the Commission has the power to conduct investigations, study issues, and gather information as directed by the General Assembly. The Commission provides in-depth research on a variety of topics, crafts recommendations to improve public policy and statutory law, and works closely with legislators and their staff.

A Commission study may involve the appointment of a legislative task force, composed of a specified number of legislators from the House of Representatives or the Senate, or both, as set forth in the enabling statute or resolution. In addition to following the progress of a particular study, the principal role of a task force is to determine whether to authorize the publication of any report resulting from the study and the introduction of any proposed legislation contained in the report. However, task force authorization does not necessarily reflect endorsement of all the findings and recommendations contained in a report.

Some studies involve an appointed advisory committee of professionals or interested parties from across the Commonwealth with expertise in a particular topic; others are managed exclusively by Commission staff with the informal involvement of representatives of those entities that can provide insight and information regarding the particular topic. When a study involves an advisory committee, the Commission seeks consensus among the members.² Although an advisory committee member may represent a particular department, agency, association, or group, such representation does not necessarily reflect the endorsement of the department, agency, association, or group of all the findings and recommendations contained in a study report.

¹ Act of July 1, 1937 (P.L.2460, No.459); 46 P.S. §§ 65 – 69.

² Consensus does not necessarily reflect unanimity among the advisory committee members on each individual policy or legislative recommendation. At a minimum, it reflects the views of a substantial majority of the advisory committee, gained after lengthy review and discussion.

Over the years, nearly one thousand individuals from across the Commonwealth have served as members of the Commission's numerous advisory committees or have assisted the Commission with its studies. Members of advisory committees bring a wide range of knowledge and experience to deliberations involving a particular study. Individuals from countless backgrounds have contributed to the work of the Commission, such as attorneys, judges, professors and other educators, state and local officials, physicians and other health care professionals, business and community leaders, service providers, administrators and other professionals, law enforcement personnel, and concerned citizens. In addition, members of advisory committees donate their time to serve the public good; they are not compensated for their service as members. Consequently, the Commonwealth of Pennsylvania receives the financial benefit of such volunteerism, along with their shared expertise in developing statutory language and public policy recommendations to improve the law in Pennsylvania.

The Commission periodically reports its findings and recommendations, along with any proposed legislation, to the General Assembly. Certain studies have specific timelines for the publication of a report, as in the case of a discrete or timely topic; other studies, given their complex or considerable nature, are ongoing and involve the publication of periodic reports. Completion of a study, or a particular aspect of an ongoing study, generally results in the publication of a report setting forth background material, policy recommendations, and proposed legislation. However, the release of a report by the Commission does not necessarily reflect the endorsement by the members of the Executive Committee, or the Chair or Vice-Chair of the Commission, of all the findings, recommendations, or conclusions contained in the report. A report containing proposed legislation may also contain official comments, which may be used in determining the intent of the General Assembly.³

Since its inception, the Commission has published more than 350 reports on a sweeping range of topics, including administrative law and procedure; agriculture; athletics and sports; banks and banking; commerce and trade; the commercial code; crimes and offenses; decedents, estates, and fiduciaries; detectives and private police; domestic relations; education; elections; eminent domain; environmental resources; escheats; fish; forests, waters, and state parks; game; health and safety; historical sites and museums; insolvency and assignments; insurance; the judiciary and judicial procedure; labor; law and justice; the legislature; liquor; mechanics' liens; mental health; military affairs; mines and mining; municipalities; prisons and parole; procurement; state-licensed professions and occupations; public utilities; public welfare; real and personal property; state government; taxation and fiscal affairs; transportation; vehicles; and workers' compensation.

Following the completion of a report, subsequent action on the part of the Commission may be required, and, as necessary, the Commission will draft legislation and statutory amendments, update research, track legislation through the legislative process, attend hearings, and answer questions from legislators, legislative staff, interest groups, and constituents.

³ 1 Pa.C.S. § 1939 ("The comments or report of the commission . . . which drafted a statute may be consulted in the construction or application of the original provisions of the statute if such comments or report were published or otherwise generally available prior to the consideration of the statute by the General Assembly").



General Assembly of the Commonwealth of Pennsylvania

JOINT STATE GOVERNMENT COMMISSION

Room 108 – Finance Building

Harrisburg, Pa 17120

717-787-4397 Fax 717-783-9380

REP. FLORINDO J. FABRIZIO

Chairman

July 2018

SEN. JOHN C. RAFFERTY, JR.

Vice Chairman

To the Members of the General Assembly of Pennsylvania:

EXECUTIVE COMMITTEE

Senate Members:

JOSEPH B. SCARNATI, III

President Pro Tempore

JACOB D. CORMAN, III

Majority Leader

JAY COSTA, JR.

Minority Leader

JOHN R. GORDNER

Majority Whip

ANTHONY H. WILLIAMS

Minority Whip

ROBERT B. MENSCH

Chair, Majority Caucus

WAYNE D. FONTANA

Chair, Minority Caucus

The Joint State Government Commission is pleased to announce the release of the report, *Evidence-Based Decision Making by the Commonwealth of Pennsylvania*, pursuant to Senate Resolution 294 of 2016.

This staff study presents a detailed explanation of evidence-based decision making (EBDM) and how it is fast becoming the decision making paradigm used by the Commonwealth. A comprehensive review of EBDM in Commonwealth agencies describes their programs and the EBDM processes in use.

House Members:

MICHAEL C. TURZAI

Speaker

DAVID L. REED

Majority Leader

FRANK J. DERMODY

Minority Leader

BRYAN D. CUTLER

Majority Whip

MICHAEL K. HANNA

Minority Whip

MARCY TOEPEL

Chair, Majority Caucus

DAN B. FRANKEL

Chair, Minority Caucus

The full report, along with a one-page summary, is enclosed and is also available on our website at <http://jsg.legis.state.pa.us>.

The full report is also available at <http://jsg.legis.state.pa.us/>.

Respectfully submitted,

Glenn J. Pasewicz
Executive Director

Administrative Staff:

GLENN J. PASEWICZ

Executive Director

YVONNE M. HURSH

Counsel

TABLE OF CONTENTS

EXECUTIVE SUMMARY	1
INTRODUCTION TO EMPIRICAL METHODS AND EVIDENCE-BASED DECISION MAKING	3
<i>The EBDM Process</i>	5
Studies	5
<i>Performance Measurement and Performance Evaluation</i>	10
EVIDENTIARY STANDARDS AND PERFORMANCE STANDARDS	15
Evidentiary Standards	15
Federal Government	18
<i>PBPP Performance Based Standards</i>	21
Goal	21
Performance Standard	22
Outcome Measures	22
Expected Practices	23
Protocols	23
Process Indicators	23
PROGRAM REGISTRIES	25
<i>Blueprints for Healthy Youth Development</i>	25
Overview	25
Assessment Tools	26
Standard of Evidence	26
<i>CrimeSolutions.gov</i>	28
Overview	28
Review Process	29
Standard of Evidence	30
<i>The Office of Juvenile Justice and Delinquency Prevention Model Program Guide</i>	32
Overview	32
Program Criteria	32
Review Process and Standard of Evidence	33
<i>Substance Abuse and Mental Health Services Administration National Registry of Evidence-Based Programs and Practices</i>	34
Overview	34
Review Process	36
Standard of Evidence	37

<i>The California Evidence-Based Clearinghouse for Child Welfare</i>	40
Overview	40
Review Process	40
Standard of Evidence	41
<i>The Coalition for Evidence-Based Policy</i>	43
Overview	43
<i>The Promising Practices Network</i>	44
Overview	44
<i>The What Works Clearinghouse</i>	44
Overview	44
Review Process	45
Standard of Evidence	45
Other Registries	46
PENNSYLVANIA STATE AGENCIES	47
Pew Foundation	49
Program Assessment	50
Return on Investment	51
Budget Development	51
Implementation Oversight and Outcome Monitoring	51
Targeted Evaluation	52
<i>Executive Branch Agencies</i>	54
Office of Administration (OA)	54
Office of the Budget (OB)	55
Office of Inspector General (OIG)	55
Department of Agriculture (PDA)	56
Department of Banking and Securities (DoBS)	58
Department of Community and Economic Development (DCED)	59
Department of Conservation and Natural Resources (DCNR)	60
Department of Corrections (DOC)	62
Department of Drug and Alcohol Programs (DDAP)	64
Department of Education (PDE)	66
Department of Environmental Protection (DEP)	67
Department of Health (DOH)	71
Department of Human Services (DHS)	71
Department of Labor and Industry (L&I)	74
Department of Revenue (DOR)	78
Department of State (DOS)	80
Juvenile Court Judges' Commission (JCJC)	82
Pennsylvania Board of Probation and Parole (PBPP)	86
Pennsylvania Commission on Crime and Delinquency (PCCD)	87
Pennsylvania Council on the Arts (PCA)	92
Pennsylvania Emergency Management Agency (PEMA)	95
Pennsylvania Liquor Control Board (PLCB)	96
Pennsylvania Turnpike Commission (PTC)	99

OTHER STATES	101
<i>States Engagements in Evidence-based Policy Making</i>	101
Washington	101
Utah	101
Minnesota	102
Connecticut	102
Oregon.....	102
 RECOMMENDATIONS FOR PENNSYLVANIA’S USE OF EVIDENCE-BASED DECISION MAKING	 103

EXECUTIVE SUMMARY

Evidence-Based Decision Making (EBDM) refers to decision makers' use of empirical research methods to make policy decisions based on analysis of evidence. The EBDM process begins with decision makers defining the concept of what policy outcomes are most desirable, followed by the formation of goals and objectives, data gathering, analysis, and evaluation. In the final step, decision makers may direct resources toward new or promising programs or away from programs that fail to meet expectations, depending on how well they achieve the desired outcomes.

Many agencies in Pennsylvania's state government engage in EBDM processes to greater or lesser extents. The report contains detailed descriptions of how Pennsylvania state government are engaged in EBDM through program assessment, budget development, implementation oversight, outcome monitoring, and targeted evaluation. Agencies like the Pennsylvania Board of Probation and Parole (PBPP) are fully invested in EBDM. PBPP, for example, has partnered with outside researchers to develop processes to help meet its goals and objectives. Programs are regularly monitored and evaluated for their performance against ongoing research; staff is regularly trained to apply methods and techniques that have been developed and proved to be effective. The use of research-based evidence has become an integral part of PBPP's agency-wide policy making.

Opportunities continue to develop for other state agencies to take full advantage of EBDM. National organizations, such as the Pew-MacArthur Trust's "Results First Initiative" are engaging Commonwealth agencies to share data, research, and technical assistance to enhance EBDM. Federal government initiatives, such as the U.S. Department of Education's "What Works Clearinghouse," provide support, guidance, and assistance to agencies that are using EBDM processes. By leveraging the evolving resources, EBDM is fast becoming the decision making paradigm in use by Pennsylvania state government agencies.

INTRODUCTION TO EMPIRICAL METHODS AND EVIDENCE-BASED DECISION MAKING

“We don’t admire gut decision makers for the quality of their decisions so much as for their courage in making them.”⁴

While this quote may ring true in many arenas, gut decisions are not particularly descriptive of the type of decisions people desire or expect in public policy development and implementation.

In concept, evidence-based decision making (EBDM) is not new. Long ago, farmers relied on generations of experience with weather patterns and constellations in the night skies to decide which crops to plant and when. As science developed the knowledge and technology to help predict crop success and decrease the risks involved, landowners relied less on traditional methods and more on what could be proven through scientific, statistical, mathematical analyses. Regardless of the vocabulary, the techniques, or the technology, from King Solomon to present day cabinet secretaries, decision makers have always relied on the information presented to them when deciding on a course of action.

The contemporary paradigm of EBDM evolved along with increasing computer power to collect and quickly analyze large datasets. Evidence that had once taken enormous investments of money, time, and people to analyze has been moved out from the computer lab onto desktop computers and now into handheld mobile devices. Computer technology is seemingly boundless as a widely accessible, inexpensively attainable resource for federal and state governments, even for resource-poor local governments, meaning that policy makers are able to take advantage of data and information gathered from their constituents in ways that were not feasible even a decade ago. With inexpensive software applications on smart phones, public policy researchers can assign hundreds of people to record and upload observations to cloud databases in real-time.

Nonetheless, policy makers’ enthusiasm generated by the abundance of information is easily overloaded; the ever-increasing mass of data brings an inertia that too often anchors policy makers to ineffective, inefficient decision making processes.

Here, however, policy makers need to align their processes with science. The sciences long ago developed means and methods of EBDM that provide for structured thought processes, information management, and strategic planning to achieve stated goals and help dispel ambiguities arising from information overload. There are six sequential steps in the classic scientific method:

“Science, however, is never conducted as a popularity contest, but instead advances through testable, reproducible, and falsifiable theories.” –
Michio Kaku

⁴ Leigh Buchanan and Andrew O’Connell, “A Brief History of Decision Making,” Harvard Business Review, January 2006, <https://hbr.org/2006/01/a-brief-history-of-decision-making>.

1. Form ideas and research questions

Policy makers need to understand what it is they intend to study. They need clear vision of the context in which they are working. Their ideas need to be concisely stated and the reasons for initiating the decision making process need to be established before any work can begin.

2. Develop theories and hypotheses

Policy makers need to be informed and knowledgeable about the topic they intend to study. Background research forms a foundation for the proposed study. Based on work that had been conducted previously, policy makers can develop their knowledge and understand the logic that led to the status quo. Policy makers can develop their own theories that are grounded in what is already known. From this point, policy makers next create their hypotheses about how a program or policy intervention affects what it is they intend to study.

3. Construct a research plan for data collection and analysis

EBDM relies on evidence. Evidence is constructed from information. Information, in turn, is distilled from data that have been collected and analyzed. Policy makers need to develop a plan for data collection that includes safeguards against bias and helps ensure reliability.

4. Implement the research design through the collection of data

Data collection takes a number of forms and can provide qualitative, quantitative, or a mixed-methods approach that includes both. In public policy research, qualitative and quantitative evidence inform the process. Anecdotal information about individuals affected by particular program interventions can be descriptive overlays to help policy makers conceptualize the stories behind the numbers and statistics; together they create stronger evidence when combined.

5. Analyze the data

Data analysis can take many different forms, whether statistical analysis of numerical data or interpretive analysis of qualitative data. Regardless of which methods are applied, researchers must follow the appropriate methodology in order to provide policy makers with helpful information.

6. Draw conclusions and communicate

At the end of the process, researchers combine each of the previous five steps into a comprehensive and concise report to policy makers. In turn, the policy makers apply the accumulated evidence to the EBDM process.

Contemporary policy makers are able to follow these steps, particularly with regard to items 4, 5, and 6, more effectively and efficiently than ever before. The flow of information from the stakeholders or populations affected by any particular set of policies up to decision makers and the flow of information that is relayed back to the stakeholders is rapid and easily disseminated to

broad audiences through the use of electronic media in the form of Internet sites, social media, and broadcast media.

Importantly, the availability of information on the Internet allows policy makers to review and analyze experiments that have been conducted by others without having to embark on their own experimentation. They can readily interpret the conclusions drawn by others and, provided that the others' research was valid, apply those conclusions to their own contexts.

The EBDM Process

The principles of the empirical method can be combined into the first of three steps that describe the overall EBDM process. First, policy makers identify needs for new programs or interventions and test those ideas. The second step of improving existing programs and identifying ineffective programs can occur in parallel. The third step involves expanding effective programs in ways that can help larger populations or in ways that can help groups outside of the original target population.

- 1. Test new ideas.**
- 2. Improve effective programs as needed and halt ineffective programs.**
- 3. Scale the effective programs.**

When properly followed, this process is iterative; that is, the researcher, at each step along the way, evaluates the planning and processing and amends the plan and process as necessary. Adherence to rigorous academic and scientific standards, particularly through use of the scientific method, will go a long way in helping policy makers produce evaluations that provide high quality evidence. It is particularly important to understand the concepts of external validity with regard to the third principle of scaling programs that lead to good outcomes. A program successful in Philadelphia may not work in Erie.

Decisions are based on evidence (decisions made without consideration of the context are essentially random and irrational).

Studies

Approaches to EBDM involve the use of evidence developed from previous research, conducting original research plans, or a combination of both. By taking advantage of Internet resources, a policy maker can learn from reports, journals, and information on every imaginable topic. These "third party" studies provide several advantages to the policy maker:

1. Third party studies limit the need for a government to perform its own studies, reducing the cost of EBDM.

2. Studies can show which programs have been tried and were found unsatisfactory and which programs offer the best opportunities for success.
3. Third party studies that were published in professional journals often have academic rigor and a strong methodological foundation to support the evidence they provide.⁵

These advantages need to be considered along with disadvantages of third party studies, which include:

1. Third party studies are not comprehensive across all topics that policy makers need information on.
2. Despite academic rigor and methodology, not all studies are conducted with appropriate internal and external validity. That is, not all studies are faithful to the methodology, and not all studies are generalizable outside of the study sample.⁶
3. Third party studies may superficially appear to address the topic, but may have a particular bias or editorial slant that influences decision makers' interpretations of the research.

Policy makers are certainly capable of conducting their own studies to inform EBDM, provided that they have the resources, not the least of which include the training, knowledge, and ability to follow empirical methods. A number of organizations have been created in recent years to inspire, inform, and, in some cases, compel EBDM in government. The Evidence-Based Policymaking Collaborative (EBPC), a collaborative of several organizations dedicated to EBDM, identifies two overarching goals of EBDM, "to use what we already know" to make policy decisions and "to build better knowledge" to inform future decisions.⁷ The EBPC developed four principles to be used as guidelines applied to the policymaking process, which are shown in Diagram 1:⁸

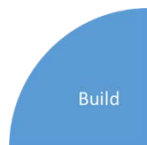
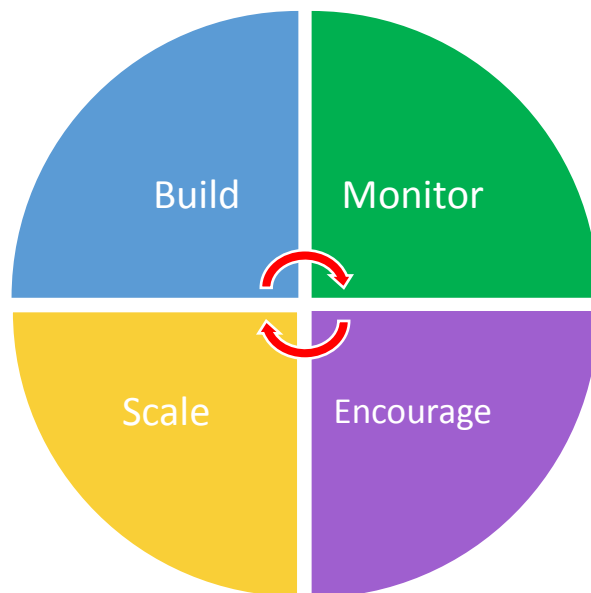
⁵ Shayne C. Kavanagh and Nate Levenson, "Running Government like a Business Scientist, The Rise of Evidence-Based Decision Making," *Government Finance Review*, August 2016 pp. 17-25, <https://www.gfoa.org/sites/default/files/GFR081618.pdf>.

⁶ Ibid.

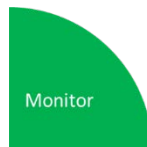
⁷ "Principles of Evidence-Based Policymaking," Evidence-Based Policymaking Collaborative, September 2016, <http://www.evidencecollaborative.org/principles-evidence-based-policymaking>. Members include the Urban Institute, the Brookings Institution, the American Enterprise Institute, and The Pew-MacArthur Results First Initiative.

⁸ Ibid.

Diagram 1
Evidence-based Policy Collaborative
Four Principles of Evidence-based Decision Making



1. Build and compile rigorous evidence about what works, including costs and benefits. The purpose of EBDM is to create and maintain a process that uses evidence in support of decision making.



2. Monitor program delivery and use impact evaluation to measure program effectiveness. This principle helps ensure that programs are operating as intended. Typically, monitoring program delivery measures “fidelity to the model,” that is, how well a given program adheres to the design it is based on. Generally, accrediting bodies monitor their constituent programs in this manner. It is important to be aware, however, that fidelity of model is not necessarily an indication that a program’s outcomes are benefiting its beneficiaries as intended.

For example, members of the advisory committee appointed by Senate Resolution 267 of 2015 to study addiction treatment services in Pennsylvania stressed that however important meeting accreditation standards are, fidelity of model is necessarily bounded by each provider’s individualized attention and treatment of each patient.⁹ Evidence-based programs provide a foundation for delivery of high quality care for patients, but the goal is to facilitate best case outcomes for patients, the accomplishment of which may diverge from the prescribed model.

Holding too closely to the parameters of evidence-based practices risks losing the individualized care needed for successful treatment. Further, members remarked that evidence-based practices are primarily applied in group therapy, whereas individualized modalities work

⁹ 2015 SR267, “Advisory Committee on Addiction Treatment Services,” JSGC meeting of July 12, 2016.

best in one-on-one sessions. In the words of an Advisory Committee member, “The purpose of evidence-based practices is not to do the model, it’s to get the best outcome.”¹⁰ In other words, it is important for clinicians and policy makers to consider not only evidence-based practices, but also *practice-based evidence*. In “Practice-Based Evidence,” Anne Swisher wrote:

In the concept of Practice-Based Evidence, the real, messy, complicated world is not controlled. Instead, real world practice is documented and measured, just as it occurs, “warts” and all. It is the process of measurement and tracking that matters, not controlling how practice is delivered. This allows us to answer a different, but no less important, question than “does X cause Y?” This question is “how does adding X...intervention alter the complex personalized system of patient Y before me?”¹¹

“The purpose of evidence-based practices is not to do the model, it’s to get the best outcome.”

Reconciling the two approaches in developing novel treatment modalities leads to debate among clinicians. Nevertheless, there is broad agreement that both evidence-based practice (the evidence having been empirically derived, documented, and peer reviewed via randomized controlled trials) and practice-based evidence (the evidence having been experienced, learned, and shared via “the real, messy complicated world” of counseling clients as they arrive at the clinic) have a role in developing and delivering the best possible treatments to patients.

Treatment is tailored to each client’s needs. For many people, initial treatment is intensive, with clients attending multiple outpatient sessions each week. As clients complete each step of their treatment plans, they can transition to less-intensive care. Clients transitioned to regular outpatient treatment meet with counselors less often and for fewer hours per week as they sustain their recovery.

To engage in outcome monitoring, that is, establishing performance measures, collecting of performance data, and reporting the information to policy makers moves the entity one step closer to EBDM. Outcome monitoring is a more involved and complex exercise than routine data collection and reporting. Outcomes are results of activities or interventions of a program, and are measured according to two different time horizons that are defined by those who are measuring the outcomes.¹² In the short term horizon, outcomes may be measured immediately after receiving the intervention, or within weeks, or within the first year after the program intervention was delivered. For example, a person receiving job training may be asked to respond to a questionnaire about the program immediately after it concludes. Follow-up observations may measure how effective the program was in helping the person find employment that he or she might not otherwise have secured without the program.

¹⁰ 2015 SR267, “Advisory Committee on Addiction Treatment Services,” JSGC meeting of July 12, 2016.

¹¹ Anne K. Swisher, “Practice-Based Evidence,” *Cardiopulmonary Physical Therapy Journal* 21.2, June 21, 2010, accessed April 4, 2017, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2879420/>.

¹² Mitchell Brown and Kathleen Hale, *Applied Research Methods in Public & Nonprofit Organizations* (Jossey-Bass, 2014), 83.



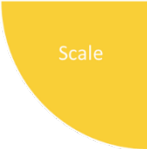
“Never confuse Motion with Action.”

Benjamin Franklin




At a long term horizon, researchers may contact the recipients of the job training program several years afterward to determine how they fared and the outcomes attributed to the program. The measures at each observation point may change, depending on the program’s intended outcomes. A public health diabetes education program may have desired long term outcomes of lower hospitalization rates, longer life expectancy, and improved quality of life; its short term outcomes may be more concerned with retention of the materials and likelihood of patients making and keeping doctor appointments.

Outcomes are different from *outputs* and are not interchangeable, though they are closely related. Outcomes are the desired consequences of outputs. Outcomes are focused on what happens after a program’s intervention. Outputs are more functional and operational. For example, measuring outputs of the diabetes education program might include a review of how costly it was to advertise, measuring the costs of education materials, counting the number of staff hours spent planning and implementing the program, tallying the number of participants, collecting their demographic data, auditing the program’s cost per participant, and similar measures. Each of these measures is vital to management of a government entity’s available resources and obligations to its constituents, and they serve as a basis for developing measures for outcomes. Measuring and evaluating outputs are not, in and of themselves, sufficient operations for EBDM.

 **Scale** 3. Use rigorous evidence to improve programs, scale what works, and redirect funds away from consistently ineffective programs. Perhaps one of the most useful results of EBDM is that decision makers are able to use research and evidence to invest funds in the most promising programs. In an environment characterized by fiscal constraints and limited resources, redirecting funds from one program to another is more easily justified when such decisions are supported by evidence.

Similarly, an initiative to scale up a successful program is made easier and more convincing when evidence supports such a move.

 **Encourage** 4. Encourage innovation and test new approaches. As each functional level of a government entity plans, implements, collects data, and evaluates its programs, the next higher level collects data from its constituent entities, evaluates the larger scope of programs, then plans and implements the next phase of programs based on what it has learned. Similarly, the next higher level does the same. As information flows upward, it may become less detailed at each level. This concept was identified and defined by John

Robert in a 1979 article in the *Harvard Business Review*, titled, “Chief Executives Define Their Own Data Needs,” which proposed “that systems used by corporate leaders ought to give them data about the key jobs the company must do well to succeed.”¹³ Although written with corporate information systems and business executives in mind, the same paradigm applies to government executives and can be deconstructed to more basic concepts. The reporting system, whether digital IT, composed of staff in an organizational chart, or a combination of both, must filter information so that the entity’s decision makers are seeing what they need to ensure that key responsibilities are being effectively managed to fulfil the agency’s obligations to its constituents.

Performance Measurement and Performance Evaluation

The next step in identifying concepts of EBDM is to understand that performance measurement and performance evaluation are not one in the same process.

Performance Measurement. Performance measurement is an ongoing process that describes how well a program functions in conducting its mandate as based on its theory of change.¹⁴ A theory of change is composed of a problem, practical solutions to the problem (i.e. programs and interventions), and the expected outcome when the solutions are applied. For example, if Pennsylvania were faced with high rates of diabetes, a theory of change could consist of:

1. Undiagnosed diabetes
2. Public education to encourage vulnerable populations to have their A1-c checked
3. Improved health for diabetes patients

An annual performance measurement of the diabetes program could consist of estimating the number of undiagnosed diabetes cases in Pennsylvania, the cost of the diabetes education program, the number of people reached by the education program, the number of people who have their A1-c checked, etc. An influenza vaccination program’s performance measurement might include counting how many doses of vaccine were administered during the flu season, how many public health workers were employed, which kind of vaccines were administered, the efficiency of the distribution system, etc. These data provide valuable information to decision makers who need to know whether the program is meeting its intended objective, that is, to administer flu vaccines to as many of the targeted demographic as possible for the lowest cost.

¹³ Leigh Buchanan and Andrew O’Connell, “A Brief History of Decision Making,” *Harvard Business Review*, January 2006, <https://hbr.org/2006/01/a-brief-history-of-decision-making>.

¹⁴Peter A Tatian, “Performance Measurement to Evaluation” *Metropolitan Housing and Communities Policy Center, Urban Institute*, (March 2016), https://www.urban.org/sites/default/files/publication/78571/2000555-performance-measurement-to-evaluation-march-2016-update_1.pdf.

A slightly more sophisticated approach would have researchers conduct a pre-post analysis, that is, researchers would collect program measure data prior to initiation of the program and then compare those data to data collected after a pre-determined period of time. With these results in hand, researchers could then determine fidelity of model—how well the program adhered to its plan and whether it achieved its objectives. Annual performance measurements of ongoing programs are essential parts of the Commonwealth’s Governor’s Executive Budget that is prepared each year. These measurements may be considered to be *formative*, that is, they measure how well a program is designed and how well it functions according to its design. A study of planning and process measures a program’s “fidelity of model.” Performance evaluations are *summative*, that is, they are used to analyze the impact that a program had on its intended target after it has delivered its services.

In more detail, a formative study may seek to identify and understand the goals and objectives of a program, the population the program intends to serve, whether the intervention is appropriate, the program’s expected impact, the evidence supporting this impact, the adequacy of the program’s resources, and the feasibility of the program’s time schedule.¹⁵

Different methods can be used to gather and analyze data in answering questions about program performance. Researchers, utilizing appropriate methodology, can analyze data and information from databases, audits, or fiscal records. They can conduct surveys, make observations, conduct interviews, and facilitate focus groups. Both quantitative and qualitative data can be analyzed through the use of appropriate levels of measurement to derive descriptive statistics and conduct multivariate analyses. The researchers’ conclusions ought to shed light on how well a given program achieves its intended goals and objective and reaches its intended population.

“Performance measurement is responsive and adaptive; evaluation answers a predetermined set of questions. While performance measurement is intended to answer questions about a program’s execution, these questions are not fixed and can change as the program evolves.”
— Peter A. Tatian, “*Performance Measurement to Evaluation.*”

Performance Evaluation. Performance evaluation, on the other hand, is a more comprehensive study of how well a program’s outcomes are serving its intended demographic. In a fundamental sense, a performance evaluation is a retrospective study that answers the question, “Did it work?” The summative performance evaluation attempts to measure the impact that the program intervention had with regard to the theory of change that guided the design of the program and was established prior to implementation. Performance evaluations focus on internal validity (that is, are we measuring what we think we are measuring?) and external validity (based on our experiment, can we generalize what we have learned? Can this program be scaled up to serve other, larger populations?). These studies are experimental in nature, going beyond the data collection

¹⁵ Peter A Tatian, “Performance Measurement to Evaluation” *Metropolitan Housing and Communities Policy Center, Urban Institute*, (March 2016), https://www.urban.org/sites/default/files/publication/78571/2000555-performance-measurement-to-evaluation-march-2016-update_1.pdf.

and analyses that conclude performance measurements, and extending into an assessment of counterfactuals, that is an assessment of what was likely to have happened to the intended population had the program not been implemented. Consideration of counterfactuals and interventions includes analyses of the null hypothesis and whether researchers can conclude whether the intervention had no effect whatsoever on the target population. In other words, unemployed people who were provided with job training would have found jobs without those job training services, or people with undiagnosed diabetes would have gotten their A1-c checked regardless of whether they were presented with diabetes education provided through a given program being evaluated. A properly conducted performance evaluation can be very valuable in providing evidence for policy makers.

A performance evaluation of a flu vaccine program could, in addition to measuring parameters outlined in a performance measurement, question the effectiveness of the vaccines for preventing serious flu infections, conduct a cost-benefit analysis of the cost per dose against the cost saved by reducing lost productivity to the economy.

Performance evaluations typically involve experiments, where one sample group receives the intervention while another does not. The outcomes of the two groups are then compared to determine whether the intervention was successful. A true experiment involves a random controlled trial. A random sample of people from the program's targeted demographic is chosen to receive the intervention (job training, child care assistance, diabetes education, for example) while another group of people is randomly selected to not receive the intervention. Afterward, the outcomes for both groups are compared. Because it is often difficult to conduct true experiments, public policy researchers may choose to conduct a quasi-experimental, or comparison, study. In a quasi-experimental study, a sample group receives the program intervention. The researchers compare changes in the treatment group to changes in groups that most closely resemble the treatment group prior to the intervention.

When performance evaluations maintain rigorous scientific standards, the conclusions drawn can provide reliable and useful evidence for policy makers.

Performance measurement and performance evaluation reside on a continuum where both provide policy makers with decision tools and evidence. Performance measurement should be conducted at regular intervals throughout a program's existence, as they give real-time measures of how well a program is functioning according to its plan. Performance evaluations can be conducted independent of the performance measurements for the purposes of drawing conclusions about the program's outcomes beyond whether it exhibits fidelity of model. An example of this can be found in substance abuse counseling.

All program planning should be iterative. That is, initial plans should be drafted and refined as policy makers go through the process of designing a program. The program's theory of change should include well-defined expectations about outcomes. To establish internal validity, the researchers need to know what they are measuring and that they are indeed measuring what they think they are measuring. After necessary adjustments to the plan have been incorporated, the program can be initiated. Performance measurement should begin collecting data and information from the moment of implementation. The data should be used to show how well the program

functions, and to further refine how the program is delivered and maintained. Concurrently, periodic performance evaluations should be conducted to determine how the program is affecting the outcomes of the intended population.

Policy makers should use the evidence harvested from the measurements and evaluations to continue to refine the program and make decisions about resource allocation relative to other public programs.

After the program has been established and functioning for an appropriate amount of time, a summative evaluation can be used to address fundamental questions about the existence of the program. Evidence can be used to make decisions about the “applicability, scalability, and replicability” of the program.¹⁶

¹⁶ Peter A Tatian, “Performance Measurement to Evaluation” Metropolitan Housing and Communities Policy Center, Urban Institute, (March 2016), https://www.urban.org/sites/default/files/publication/78571/2000555-performance-measurement-to-evaluation-march-2016-update_1.pdf.

EVIDENTIARY STANDARDS AND PERFORMANCE STANDARDS

Evidentiary Standards

Evidentiary standards, the criteria by which data are measured and the parameters by which evidence is determined to be reliable, vary across disciplines, largely depending on the topic under consideration. Further, evidentiary standards are variable depending on the capabilities of a program and on the desired and resulting outcomes. An example of variability in accepted outcomes is found in public health policies related to exposure to lead in the environment. There is no minimally safe level of lead; lead exposure is unsafe no matter how low the level of exposure may be. However, achieving zero exposure is not practical given current technology and infrastructure. Therefore, public health authorities establish reference values for lead exposure based on blood lead level measurements in children ages 1-5 years.¹⁷

Broadly speaking, there are different levels of evidentiary standards. At the lowest level, nonexperimental research gathers evidence from anecdotal, testimonial, and personal experiences that are conveyed to policy makers. Evidence gathered from nonexperimental research can include surveys, interviews, agency records and archives, results from focus groups, and case studies. Table 1 lists the different methods of data gathering in nonexperimental research. Nonexperimental research yields descriptions of populations and communities, and helps policy makers understand problems and issues. Typically categorized as qualitative research, with firm ground rules and plans established, this type of research can contribute significant knowledge to policy makers. There is no doubt that non-experimental/qualitative research is critically important during early stages of policy planning, and frequently germinates the impetus to improve existing and develop new policies.

¹⁷ Oscar Tarragó, Mary Jean Brown, "Lead Toxicity Case Studies in Environmental Medicine," *Agency for Toxic Substances and Disease Registry, Environmental Health and Medicine Education*, Accessed May 9, 2018, <https://www.atsdr.cdc.gov/csem/csem.asp?csem=34&po=8>.

Table 1
Approaches to Nonexperimental Data Collection

Data Collection Approach	Overview	Strengths and Weaknesses
Direct Observation	Passive or active observation of phenomenon	(+) in-depth (-) highly subjective and often unstructured
Focus Groups	Small-group moderated conversation	(+) interactive and in-depth (-) reliability
Elite or Key Informant Interviews	Conversation with experts or key informants, with varying degrees of structure	(+) in-depth (-) reliability, validity
Case Studies	In-depth examination of a phenomenon or place, utilizing multiple data forms	(+) in-depth, causality (-) generalizability, reliability
Content Analysis	Systematic analysis of recorded materials	(+) in-depth (-) reliability
Surveys	Oral or written collection of information about beliefs and behaviors; demographics	(+) in-depth (-) validity, reliability, subjectiveness
Secondary Data Analysis	New or repeat analysis of data others have previously collected	(+) minimal effort/cost to collect

Source: Mitchell Brown and Kathleen Hale, *Applied Research Methods in Public and Nonprofit Organizations*, Josey-Bass, (San Francisco, CA: 2014).

While nonexperimental evidence can be an indication of how effective a particular program is, it does not account for external influences and does not control for internal and external validity.

Measures of correlation between a program’s inputs and observed outcomes can show connections between policy and its purported results. It is important here to recall the axiom, “Correlation is not causation.” A strong correlation between two variables, e.g. increased employment and job training, might indicate that unemployment decreases when job training programs are provided, and vice versa, but does not show that job training causes unemployment rates to drop.

Higher levels of evidentiary standards are met by research that incorporates experimental studies. These studies, which include pre-tests and post-tests and experimental and control groups, are designed to meet the rigors of statistical analysis. There are different levels of confidence among types of experimental studies, although each involves comparison between groups to test for statistical significance, that is, to test whether a program is likely to have achieved in its intended outcomes.

A quasi-experimental test records data and observations of outcomes exhibited by a group of people before and after they receive the intervention. The data and observations are compared with those made of a similar group that did not receive the intervention. The analyses may indicate correlation, significance, and causation, but does not control for unmeasured variables. A quasi-experimental test may be used when ethical or practical considerations prevent the use of random sampling. For example, if a policy maker wanted to determine the effects of a new methadone maintenance treatment program for people with substance use disorders, it would be unethical to randomly assign people to different methadone maintenance treatment programs without considering their individual needs.

The highest level of evidence comes from experiments that are constructed around random samples of experimental and control groups, and can be even more reliable when multiple experiments are conducted.

The highest level of evidentiary standard is met when research adheres to the scientific method. The experiments that generate data and evidence begin with test subjects randomly assigned to treatment and control groups. Pre-tests and post-tests of both sample groups are administered to establish baselines and possible changes caused by the program intervention. Tests of statistical significance are conducted to determine differences between pre-test and post-tests, and to determine whether there are differences between the treatment and control groups.

Properly conducted experiments also control for a number of biases that can otherwise affect the results and lead to wrong conclusions. Astute researchers will recognize how to control for selection bias, endogenous change, external events, contamination between the treatment and control groups, placebo effects and Hawthorne effects.¹⁸ The objective of taking these measures is to establish internal validity, that is, to show a causal relationship exists between the program intervention and the observed outcomes.

Having established internal validity, researchers also want to determine external validity, that is, how generalizable the program intervention is, how applicable it is to groups other than the treatment group. For example, policy makers will want to know if a program to eliminate food deserts in one particular urban setting can be successful in other urban settings, and if not, how the program needs to be modified.

The federal government provides dozens of resources that list research-based and evidence-based guidance for public policy programs. Experimental research exists in every federal department, across all disciplines, from the Department of Defense's testing of cold weather military gear to the U.S. Forest Service's tests of micrometeorology and grass fires. Social welfare programs receive rigorous testing, as well. The Substance Abuse and Mental Health Services Administration (SAMHSA) maintains the National Registry of Evidence-based Programs and Practices (NREPP), which reviews and rates programs for mental health and substance abuse interventions.

¹⁸ "Hawthorne effect" refers to the phenomenon that may occur when test subjects behave differently because they are aware that they are subjects of a study.

State governments' departments share data with their federal counterparts, which provide both state and federal policy makers with the data and evidence necessary to make program decisions to benefit their constituents. Further, state governments often require that programs that have been studied, vetted, and approved by federal agencies be used, as applicable in their own states. For example, the Pennsylvania Department of Drug and Alcohol Programs will require that Pennsylvania drug and alcohol treatment centers comply with guidelines set by SAMHA when using an approved set of programs. Similarly, the Pennsylvania Department of Health provides funding for public health education programs (such as for diabetes awareness) that have met with positive reviews by the federal Department of Health and Human Services.

Federal government

The federal government's Commission on Evidence-Based Policy Making (CEP) released a comprehensive report that discussed the role of EBDM at the federal level and developed a number of recommendations for federal agencies.¹⁹ CEP found that the federal government could utilize existing data provided that privacy and legal protections were strengthened, and transparency and accountability were increased.²⁰ Importantly, CEP stated that:

Traditionally, increasing access to confidential data presumed significantly increasing privacy risk. The Commission rejects that idea. The Commission believes there are steps that can be taken to improve data security and privacy protections beyond what exists today, while increasing the production of evidence.²¹

To address those widespread concerns, CEP recommended that privacy protections be modernized, that the National Secure Data Service be established, and that the federal government's evidence-building capacity is strengthened.²²

Building the capacity relies on several steps, according to CEP.

1. Identify or establish a Chief Evaluation Officer in each department to coordinate evaluation and policy research and to collaborate with other evidence-building functions within Federal departments.
2. Develop learning agendas in Federal departments to support the generation and use of evidence to address the range of policymakers' questions.
3. Improve coordination of government-wide evidence building by directing the Office of Management and Budget to facilitate cross-government coordination, and consider how a greater commitment to foundational information policy responsibilities can be achieved, including through any consolidation or reorganization at the Office of Management and Budget that may be necessary.

¹⁹ "The Promise Of Evidence-Based Policymaking," *Commission on Evidence-Based Policymaking*, Accessed April 9, 2018, (Washington, D.C.: September 2017), <https://www.cep.gov/cep-final-report.html>.

²⁰ Ibid.

²¹ Ibid.

²² Ibid.

4. Align administrative processes with evidence-building activities, including those relating to the approval of information collections and the procurement of services for evidence building.
5. Ensure that sufficient resources to support evidence-building activities are available, including resources to support implementation of the Commission’s recommendations.²³

The federal government established the What Works Clearinghouse (WWC) in 2002 within the federal Department of Education through its Institute of Education Sciences.²⁴

The WWC: Reviews the existing research on different programs, products, practices, and policies in education. Our goal is to provide educators with the information they need to make evidence-based decisions. We focus on the results from high-quality research to answer the question “What works in education?”

The WWC conducts “study reviews,”

An examination of the effect of an intervention on a particular sample and set of outcomes. In most cases, the results of a study are presented in a single article or manuscript. A study may also present results over multiple manuscripts, such as an examination of a beginning reading program that analyzes both immediate and long-term effects of an intervention on a common sample that is reported on in manuscripts produced over several years. In contrast, a single manuscript may contain multiple studies, such as an examination of a dropout prevention program analyzed separately in three different cities.²⁵

Topics are defined as:

An outcome, population, or type of intervention studied by the WWC. Topics are defined by intended outcome (such as improving reading skills), intended population (such as English language learners), and types of interventions (such as education technology) that may improve outcomes for students.²⁶

Topics include:²⁷

- Adolescent Literacy
- Beginning Reading
- Character Education
- Charter Schools
- Children and Students With an Autism Spectrum Disorder
- Children and Students With Intellectual Disability
- Children Identified With or at Risk for an Emotional Disturbance
- Dropout Prevention
- Early Childhood Education

²³ Ibid.

²⁴ U.S. Department of Education, Institute of Education Services, and What Works Clearinghouse, “Who We Are,” accessed April 16, 2018, <https://ies.ed.gov/ncee/wwc/WhoWeAre/>.

²⁵ Ibid.

²⁶ Ibid.

²⁷ Ibid.

- Early Childhood Education for Children with Disabilities
- English Language Learners
- Interventions for Students in Developmental Education
- Postsecondary Education
- Primary Mathematics
- Science
- Secondary Mathematics
- Students with Learning Disabilities
- Supporting Postsecondary Success
- Teacher Training, Evaluation, and Compensation
- Transition to College

Each topic is overseen by review teams, each of which includes content expert, methodological expert, and review staff.²⁸

“Intervention Reports” are written by the respective review teams to evaluate and convey findings of studies that are submitted to the WWC. Each Intervention Report is:

A summary of findings of the highest-quality research on a given program, practice, or policy in education. The WWC searches for all research studies on an intervention, reviews each against evidence standards, and summarizes the findings of those that meet standards.²⁹

WWC review teams also prepare “Practice Guides,” which are characterized as:

A WWC publication that presents recommendations for educators to address challenges in their classrooms and schools. They are based on reviews of research, the experiences of practitioners, and the expert opinions of a panel of nationally recognized experts.³⁰

²⁸ Ibid.

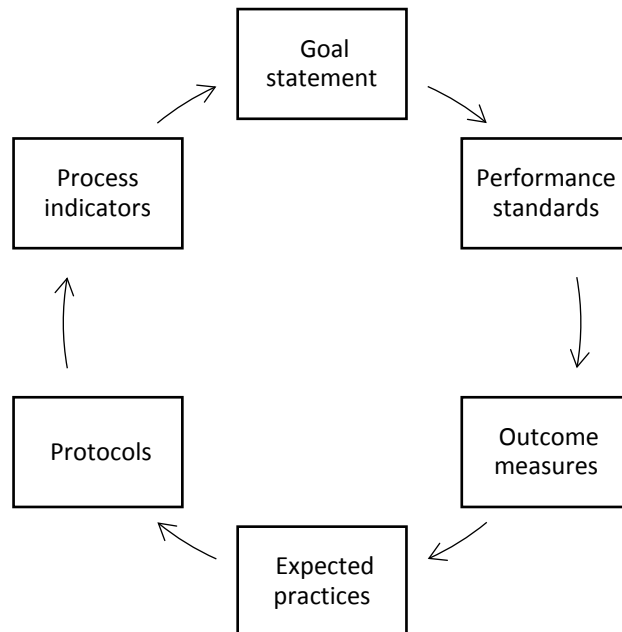
²⁹ Ibid.

³⁰ Ibid.

PBPP Performance Based Standards

In January 2011 the Pennsylvania Board of Probation and Parole (PBPP) adopted the “Performance Based Standards for Adult Probation and Parole Field Services, Fourth Edition,” in order to maintain accreditation with the American Correctional Association (ACA).³¹ The standards and accepted practices developed by the ACA are designed to “enable administrators and practitioners to not only monitor activities but also to measure over time the outcomes of their efforts.”³²

However, an attentive reader will quickly recognize that the six elements comprising the standards are generally applicable to any process of EBDM.³³



Note that the most significant are the goal statement and performance standards. No successful EBDM process begins without engaging in the practice of stating explicitly the goals of the program and the performance standards to which it will be held. However, it also is imperative that policy makers consider adjustments to goals and standards as they become necessary. Indeed, conceptualizing, developing, and implementing a mechanism for evaluating programs is itself a process subject to monitoring, evaluation, and modification.

Goal

The Goal is an overall statement of what policy makers seek to achieve through a program or policy. It can be difficult to make a clear, concise, and comprehensive goal statement because

³¹ American Correctional Association, *Performance Based Standards for Adult Probation and Parole Field Services, Fourth Edition*, (Alexandria, VA: 2010).

³² Ibid. xvi.

³³ Ibid.

the vision becomes clouded by questions of how and why. The former is better left for later discussions during the planning and development of EBDM. The latter is answered during the next stage of the process, the development of performance standards.

Performance Standard

A Performance Standard is a statement that clearly defines a required or essential condition to be achieved and maintained. A Performance Standard describes a “state of being,” and does not describe the activities of practices that might be necessary to achieve compliance. ACA gives a detailed description of how the development of performance standards is not an easy task.

In the beginning, the drafters of the new standards found it difficult to articulate clear and concise standards. The closer a draft standard came to meeting the definition of a performance-based standard, the simpler it seemed to appear. In drafting the new performance-based standards, the committee was constantly fighting the urge to describe an activity rather than to identify the overarching purpose for the activity. During many of these working group meetings, it was common for the proposer of a standard to be met with the response, “Why?” While frustrating, by continuing to ask the why question, the drafters were able to identify the basic statements of conditions that must be defined through performance standards.³⁴

In fact, the quality of the entire EBDM process relies on the strength and robustness of the Performance Standards. How well a program meets its obligation to the Performance Standards is determined by the outcome measures.

Outcome Measures

ACA defines outcome measures as quantifiable (measurable) events, occurrences, conditions, behaviors, or attitudes that demonstrate the extent to which the condition described in the corresponding performance standard has been achieved.³⁵

Outcomes are the results of program activities; they are the consequences of interventions. Outcomes are distinct from the activities themselves. In some cases, program activities are referred to as outputs.³⁶ Outputs lead to outcomes. A public health program to distribute flu vaccines will conduct various activities to supply a certain number of vaccines to medical centers. The output of the distribution activities can be measured by tallying the number of vaccines at each medical center. Outputs are direct measures of the activities. Outcomes, on the other hand, are measurements of the consequences of the outputs. In the flu vaccine example, an outcome could be a measure of how effective the program was at lowering reported flu cases.

Outcomes are the measurable expression of performance standards. In turn, the expected practices are the application of the performance standards.

³⁴ Ibid., xx

³⁵ Ibid.

³⁶ Mitchell Brown and Kathleen Hale, *Applied Research Methods in Public & Nonprofit Organizations* (Jossey-Bass, 2014), 83.

Expected Practices

Expected Practices are “actions and activities that, if implemented properly, will produce the desired outcome—achievement of the condition described in the standard.”³⁷ Expected practices have to follow protocols in order to maintain their fidelity to the program design.

Protocols

Protocols are the instructions that are written to direct the proper implementation of the Expected Practices.

Process Indicators

Process Indicators are used to monitor activities and practices. It is important to remember that, “Process indicators are not an end in and of themselves—they just tell if the expected practices are being implemented.”³⁸ The process indicators are the measures that communicate to decision makers whether a particular activity or intervention is being conducted, and whether it is maintaining fidelity to the protocols and expected practices. A process indicator would be a count of the number of influenza vaccines delivered to medical centers, or a measure of the wait-time for residents calling PennDOT for vehicle registrations.

Process indicators are not the same as Outcome measures; process indicators keep track of the steps that lead to expected outcomes. Outcome measures are used to evaluate how effective the expected practices are at achieving the performance standards, and ultimately the goals.

The ACA’s emphasis on the iterative nature of EBDM is noted throughout its discussion of the six components.

But as the field continues to learn from experience, it is predicted, and even hoped, that the expected practices that are prescribed to achieve compliance with standards will continue to evolve.³⁹

An important obligation of the people involved in EBDM is that they continue to observe their process to learn from the experience and apply those lessons as the process unfolds. The EBDM process is not linear. It does not start on a certain date and end on a certain date in a discrete flow of operations. Work is reviewed and refined by decision makers “at each step and across each step.”⁴⁰ Real-time revisions to the EBDM process do not weaken the results, but rather strengthen their integrity.

³⁷ Ibid., xx

³⁸ Ibid. xxi.

³⁹ Ibid.

⁴⁰ Mitchell Brown and Kathleen Hale, *Applied Research Methods in Public & Nonprofit Organizations* (Jossey-Bass, 2014), 22.

PROGRAM REGISTRIES

Because unproven programs can not only waste scarce resources, but can also do harm, many organizations have assembled registries of programs that meet their standards to qualify as evidence-based in order to prevent such outcomes. These registries identify, recommend, and disseminate programs that, based on scientific evaluations, have strong evidence of effectiveness. The registries serve as resources for governmental agencies, schools, foundations, and community organizations trying to make informed decisions about programs they may want to implement. This section describes a few of these program registries, which were identified in Senate Resolution No. 294.

Blueprints for Healthy Youth Development

Overview

Blueprints for Healthy Youth Development (Blueprints) is hosted by the Center for the Study and Prevention of Violence (CSPV), at the Institute of Behavior Science, University of Colorado Boulder, and funding is provided by the Annie E. Casey Foundation.⁴¹ It provides a registry of evidence-based positive youth development programs designed to promote the health and well-being of children and teens.⁴² The programs are family, school, and community-based and target all levels of need, ranging from broad prevention programs that promote positive behaviors while decreasing negative behaviors, to highly-targeted programs for at-risk children and troubled teens.⁴³ Programs are identified based upon an initial review of their evaluation evidence by CSPV and a final review and recommendation from an advisory board consisting of seven experts in the field of positive youth development.⁴⁴ More than 1,400 programs have been reviewed, but fewer than five percent of them have been designated as model and promising programs.⁴⁵

⁴¹“Blueprints for Healthy Youth Development,” *blueprints programs*, accessed May 2018, <http://www.blueprintsprograms.com/about>.

⁴² Ibid.

⁴³ Ibid.

⁴⁴ Ibid.

⁴⁵ Ibid.

Assessment Tools

Blueprints has identified several surveys that facilitate matching children’s strengths and needs to specific programs.⁴⁶ One survey used is the Evidence2Success Childhood and Youth Experience Survey, a product of the Annie E. Casey Foundation that captures information on whether kids are meeting developmental milestones and where they are likely to face challenges in the future.⁴⁷

Another survey used is the Communities That Care (CTC) Youth Survey, which is a community needs assessment tool that provides community leaders with accurate estimates of risk and protective factors affecting academic performance and positive youth development, as well as alcohol and substance use, school dropout, sexual risk behaviors, delinquency and violence.⁴⁸ The CTC Youth Survey is in the public domain.⁴⁹

A third survey used is CSPV’s own Safe Communities Safe Schools (SCSS) Initiative, which offers an online school climate survey for students, staff, and parents.⁵⁰ Schools receive a comprehensive report about their strengths and challenges regarding school climate and risk and protective factors for violence, and administrators receive a report about the evidence-based programs and strategies available to address common school challenges.⁵¹

Standard of Evidence

Blueprint-endorsed programs have been reviewed by an independent panel of evaluation experts and determined to meet a clear set of scientific standards.⁵² Programs meeting these standards have demonstrated at least some effectiveness for changing targeted behavior and developmental outcomes, and are rated as either “Promising” or “Model.”⁵³ “Promising” programs meet the minimum standard of effectiveness, while “Model” programs meet a higher standard and provide greater confidence in their program capacity to change behavior and developmental outcomes.⁵⁴

“Promising” programs meet the following standards:

- 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.

⁴⁶ “Blueprints for Healthy Youth Development,” *blueprints programs*, accessed May 2018, <http://www.blueprintsprograms.com/assess-needs>.

⁴⁷ Ibid.

⁴⁸ Ibid.

⁴⁹ Ibid.

⁵⁰ Ibid.

⁵¹ “Blueprints for Healthy Youth Development,” *Blueprints Programs*, accessed May 2018, <http://www.blueprintsprograms.com/assess-needs>.

⁵² “Blueprints for Healthy Youth Development,” *Blueprints Programs*, accessed May 2018, <http://www.blueprintsprograms.com/criteria>.

⁵³ Ibid.

⁵⁴ Ibid.

- Evaluation quality: The evaluation trials produce valid and reliable findings. This requires a minimum of (a) one high quality randomized control trial 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.
- Dissemination readiness: The program is currently 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. European programs have not undergone the Blueprints certification process to determine dissemination readiness.⁵⁵

“Model” programs meet these additional standards:

- Evaluation Quality: A minimum of (a) two high quality randomized control trials or (b) one high quality randomized control trial plus one high quality quasi-experimental evaluation.
- Positive intervention impact is sustained for a minimum of 12 months after the program intervention ends.⁵⁶

“Model Plus” programs meet one additional standard: independent replication--in at least one high quality study demonstrating desired outcomes, authorship, data collection, and analysis has been conducted by a researcher who is neither a current nor past member of the program developer’s research team and who has no financial interest in the program.⁵⁷

Blueprints created seals that help people easily identify programs that meet the various standards, which are reproduced in Figure 1.

⁵⁵ Ibid.

⁵⁶ Ibid.

⁵⁷ Ibid.

Figure 1
Program Seals



Source: <http://www.blueprintsprograms.com/criteria>

Furthermore, Blueprints has created guidelines for program developers who recognize the value of certification by Blueprints and who wish to get their program approved or recommended.⁵⁸

CrimeSolutions.gov

Overview

The National Institute of Justice (NIJ) is a bureau within the Office of Justice Programs (OJP), within the U.S. Department of Justice.⁵⁹ NIJ is the research, development, and evaluation agency of the U.S. Department of Justice.⁶⁰ CrimeSolutions.gov is an initiative of OJP that is operated by NIJ, and is intended to be a central, reliable resource to help practitioners and policymakers understand what works in justice-related programs and practices.⁶¹

CrimeSolutions.gov includes a web-based clearinghouse of programs and practices, and a process for identifying and rating those programs and practices.⁶² The programs and practices presented in the clearinghouse have undergone rigorous evaluations and meta-analyses, assessing the strength of the evidence about whether these programs achieve criminal justice, juvenile justice, and crime victim services outcomes in order to inform practitioners and policy makers.⁶³

⁵⁸ "Blueprints for Healthy Youth Development," blueprints programs, accessed May 2018, <http://www.blueprintsprograms.com/guidelines>.

⁵⁹ "FAQs," Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/faqs.aspx>.

⁶⁰ "Welcome," National Institute of Justice, accessed May 2018, <https://www.nij.gov/about/Pages/welcome.aspx>.

⁶¹ "FAQs," Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/faqs.aspx>.

⁶² "About," Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/About.aspx>.

⁶³ Ibid.

Each screened program and practice is reviewed by two certified reviewers using objective scoring instruments, and ratings are assigned based on the consensus score, which is subject to a documented dispute resolution process when necessary.⁶⁴

Review Process

CrimeSolutions.gov defines a program as a specified set of activities combined according to precise guidance in order to achieve a specific purpose.⁶⁵ To be included on CrimeSolutions.gov, programs undergo an eight-step review and evidence-rating process as follows:

1. Preliminary program identification.
2. Initial program screening.
3. Literature search.
4. Initial evidence screening.
5. Selection of evidence base.
6. Expert review.
7. Study classification.
8. Program evidence rating.⁶⁶

CrimeSolutions.gov defines a practice as a general category of programs, strategies, or procedures that share similar characteristics with regard to the matters they address and how they do it.⁶⁷ To be included on CrimeSolutions.gov, practices undergo a seven-step review and evidence-rating process as follows:

1. Preliminary practice identification.
2. Initial screening.
3. Practice definition and literature search.
4. Initial evidence screening.
5. Expert review.
6. Outcome classification.

⁶⁴ Ibid.

⁶⁵ “FAQs,” Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/faqs.aspx>.

⁶⁶ “Program Review and Rating from Start to Finish,” Office of Justice Programs, accessed May 2018, https://www.crimesolutions.gov/about_starttofinish.aspx.

⁶⁷ “FAQs,” Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/faqs.aspx>.

7. Outcome evidence rating.⁶⁸

Reviewers are responsible for reviewing and rating the individual studies that comprise a program or practice's evidence base.⁶⁹ Reviewers are assigned on the basis of their area of expertise.⁷⁰ Reviewers are subject-matter and research methodology experts, and are trained and certified prior to reviewing studies.⁷¹ As minimum qualifications, each Study Reviewer must:

- Possess a Ph.D. (or other comparable advanced degree), preferably in a social science-related field of study such as sociology or criminal justice, but other fields of study will be considered;
- Be subject-matter experts and have substantial knowledge of research in at least one or more of the following seven topic areas: corrections and reentry, courts, crime and crime prevention, drugs and substance abuse, juvenile justice, law enforcement, and crime victims and victimization;
- Have extensive knowledge about research methodology; and
- Be available to complete reviews up to five days per year.⁷²

Based on the study reviewers' assessment of the evidence, CrimeSolutions.gov classifies programs and practices as "Effective," "Promising," or "No Effects."⁷³ The classifications are defined as follows:

- "Effective" programs or practices have strong evidence indicating they achieve their intended outcomes;
- "Promising" programs or practices have some evidence indicating they achieve their intended outcomes; and
- "No Effects" programs or practices have strong evidence indicating that they did not achieve their intended outcomes.⁷⁴

Standard of Evidence

CrimeSolutions.gov characterizes program and practice evaluations as they sit along an evidence continuum, as depicted in Figure 2.⁷⁵ The continuum has two axes: Effectiveness, and Strength of Evidence.⁷⁶ Where a program or practice sits on the Effectiveness axis determines how well it works in achieving outcomes; where it sits on the Strength of Evidence axis determines

⁶⁸ "About practice review," Office of Justice Programs, accessed May 2018, https://www.crimesolutions.gov/about_practicereview.aspx.

⁶⁹ "CrimeSolutions.gov Researchers and Reviewers," Office of Justice Programs, accessed May 2018, https://www.crimesolutions.gov/about_researchers.aspx.

⁷⁰ Ibid.

⁷¹ Ibid.

⁷² Ibid.

⁷³ "FAQs," Office of Justice Programs, accessed May 2018, <https://www.crimesolutions.gov/faqs.aspx>.

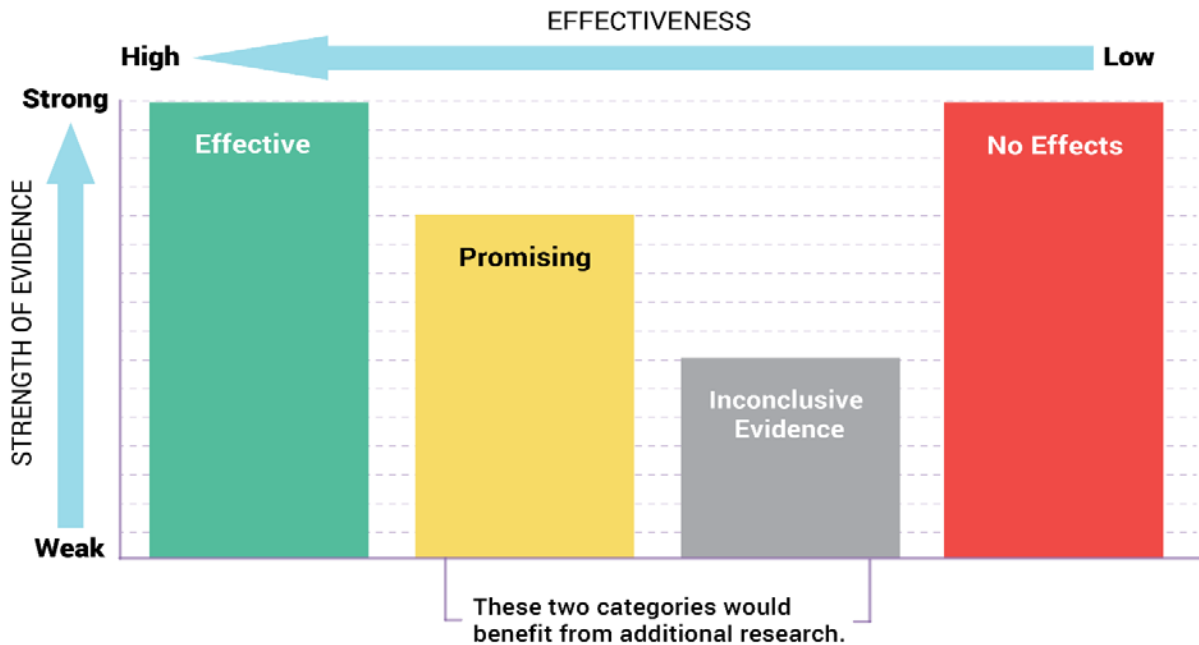
⁷⁴ Ibid.

⁷⁵ "CrimeSolutions.gov Evidence Continuum," Office of Justice Programs, accessed May 2018, https://www.crimesolutions.gov/about_evidencecontinuum.aspx.

⁷⁶ Ibid.

how confident CrimeSolutions.gov can be of that determination.⁷⁷ Effectiveness is determined by the outcomes of an evaluation in relation to the goals of the program or practice.⁷⁸ Strength of Evidence for programs is determined by the rigor and design of the outcome evaluation, and by the number of evaluations, whereas Strength of Evidence for practices is determined by the rigor and design of the studies included in the meta-analysis.⁷⁹

**Figure 2
Evidence Continuum**



Source: https://www.crimesolutions.gov/about_evidencecontinuum.aspx

On CrimeSolutions.gov, programs and practices fall into one of four categories, listed below in order of effectiveness from the continuum:

- Rated as “Effective”: Programs and practices have strong evidence to indicate they achieve criminal justice, juvenile justice, and victim services outcomes when implemented with fidelity.
- Rated as “Promising”: Programs and practices have some evidence to indicate they achieve criminal justice, juvenile justice, and victim services outcomes. Included within the “Promising” category are new, or emerging, programs for which there is some evidence of effectiveness.

⁷⁷ Ibid.

⁷⁸ Ibid.

⁷⁹ Ibid.

- “Inconclusive Evidence”: Programs and practices that made it past the initial review but, during the full review process, were determined to have inconclusive evidence for a rating to be assigned. Interventions are not categorized as inconclusive because of identified or specific weaknesses in the interventions themselves. Instead, reviewers have determined that the available evidence was inconclusive for a rating to be assigned.
- Rated as “No Effects”: Programs have strong evidence indicating that they had no effects or had harmful effects when implemented with fidelity.⁸⁰

The Office of Juvenile Justice and Delinquency Prevention Model Program Guide

Overview

Congress enacted the Juvenile Justice and Delinquency Prevention Act⁸¹ in 1974, establishing the Office of Juvenile Justice and Delinquency Prevention (OJJDP) to support local and state efforts to prevent delinquency and improve the juvenile justice system. The OJJDP developed the Model Program Guide (MPG) to provide information about evidence-based juvenile justice and youth prevention, intervention, and reentry programs.⁸²

MPG is a resource for practitioners and communities about what works, what is promising, and what does not work in juvenile justice, delinquency prevention, and child protection and safety.⁸³ MPG addresses a range of issues including child victimization, substance abuse, youth violence, mental health and trauma, and gang activity.⁸⁴ MPG provides program profiles, as well as useful literature reviews, implementation information, and links to additional resources that practitioners, policymakers, and communities can use to improve the effectiveness of interventions, enhance accountability, ensure public safety, and reduce recidivism.⁸⁵

Program Criteria

In order to be reviewed for inclusion in MPG, a program must target juveniles and meet the CrimeSolutions.gov program scope and evaluation studies screening criteria.⁸⁶ A program must aim to:

- Prevent or reduce crime, delinquency, or related problem behaviors such as aggression, gang involvement, and/or school attachment;

⁸⁰ Ibid.

⁸¹ Pub. L. No. 93-415, 42 U.S.C. § 5601 *et seq.*

⁸² “Model Programs Guide,” Office of Juvenile Justice and Delinquency Prevention, accessed May 2018, <https://www.ojjdp.gov/mpg/Home/About>.

⁸³ Ibid.

⁸⁴ Ibid.

⁸⁵ Ibid.

⁸⁶ Ibid.

- Prevent, intervene, or respond to victimization;
- Improve justice systems or processes; or
- Target an offender population or an at-risk population (that is, individuals who have the potential to become involved in the justice system).⁸⁷

There must also be existing evaluation research studies that meet the following criteria:

- The program must be evaluated with at least one randomized field experiment or a quasi-experimental research design (with a comparison condition);
- The outcomes assessed must relate to crime, delinquency, or victimization prevention, intervention, or response;
- The evaluation(s) must be published in a peer-reviewed publication or documented in a comprehensive evaluation report; and
- The date of publication must be 1980 or after.⁸⁸

Review Process and Standard of Evidence

MPG uses expert study reviewers and CrimeSolutions.gov’s program review process, scoring instrument, and evidence standards.⁸⁹ The study reviewers have extensive expertise in juvenile justice issues as well as research methodology and have completed the CrimeSolutions.gov reviewer training process.⁹⁰ Based on the study reviewers’ assessment of the evidence, programs are rated as “Effective,” “Promising,” or “No Effects.”⁹¹ The ratings are equivalent to those of CrimeSolutions.gov.⁹²

⁸⁷ Ibid.

⁸⁸ Ibid.

⁸⁹ Ibid.

⁹⁰ Ibid.

⁹¹ Ibid.

⁹² Ibid.

***Substance Abuse and Mental Health Services Administration
National Registry of Evidence-Based Programs and Practices***

SAMHSA made a significant announcement about the NREPP database in early 2018. The agency decided to overhaul vetting for evidence-based practices for inclusion in NREPP. As of December 2017, NREPP ceased accepting new additions to the database. The following information reflects SAMHSA's process as it existed prior to December 2017. No further information was available at the time this report was written.

(Source: SAMHSA, "Statement of Elinore F. McCance-Katz, MD, PhD, Assistant Secretary for Mental Health and Substance Use regarding the National Registry of Evidence-based Programs and Practices and SAMHSA's new approach to implementation of evidence-based practices (EBPs)," January 11, 2018, <https://www.samhsa.gov/newsroom/press-announcements/201801110330>. Sharon Begley, "Trump administration halts 'evidence-based' program that evaluates behavioral health therapies," *Stat*, January 10, 2018, <https://www.statnews.com/2018/01/10/trump-behavioral-health-evidence-based/>.)

Overview

The Substance Abuse and Mental Health Services Administration (SAMHSA) is the agency within the U.S. Department of Health and Human Services that leads public health efforts to advance the behavioral health of the nation, with a mission to reduce the impact of substance abuse and mental illness on America's communities.⁹³ Prior to its suspension, the National Registry of Evidence-based Programs and Practices (NREPP) was an important resource SAMHSA provided to improve access to information on evaluated interventions and to reduce the lag time between creation of scientific knowledge and its practical application in the field.⁹⁴ NREPP had been an evidence-based repository and review system designed to provide the public with reliable information on mental health and substance use interventions.⁹⁵

All interventions in the registry had met minimum requirements for review.⁹⁶ For instance, to be eligible for NREPP, an intervention must have been evaluated with at least one study that had been conducted using an experimental or quasi-experimental design.⁹⁷ Furthermore, the results of these studies had to have been published in a peer-reviewed journal or professional publication, or documented in a comprehensive evaluation report, published in 1995 or after.⁹⁸

⁹³ "About Us," The Substance Abuse and Mental Health Services Administration, accessed, May 2018, <https://www.samhsa.gov/about-us>.

⁹⁴ Ibid.

⁹⁵ Ibid.

⁹⁶ Ibid.

⁹⁷ "Review Submission," The Substance Abuse and Mental Health Services Administration, accessed, May 2018, https://nrepp.samhsa.gov/reviews_submission.aspx

⁹⁸ Ibid.

The programs' effects on individual outcomes of all interventions in the registry had been independently assessed and rated by certified reviewers.⁹⁹ Some intervention developers elected to participate in NREPP through self-nominations, while other interventions could be identified through literature searches or by SAMHSA.¹⁰⁰

NREPP assessed the research evaluating an intervention's impact on outcomes.¹⁰¹ NREPP ratings took into account the methodological rigor of evaluation studies, the size of a program's impact on an outcome, the degree to which a program was implemented as designed, and the strength of a program's conceptual framework.¹⁰²

For each intervention reviewed, NREPP published a report called a program profile.¹⁰³ Each program profile included:

- A sidebar snapshot with outcome-level ratings, program contact information, age, gender, special populations, settings, and program type;
- A brief description of the intervention, its goals, and major components;
- A presentation of the program's positive or negative impact on the outcomes reviewed;
- A description of the evaluation studies that assessed the program's effectiveness;
- Information on resources for dissemination and implementation, including available training, education, and implementation supports, as well as cost and purchasing information; and
- References.¹⁰⁴

The information and ratings included on NREPP were not static, and as additional evidence became available, the information and ratings were updated and supplemented to reflect the most current information and research.¹⁰⁵ NREPP relied on developers to provide updated information on their programs. Furthermore, NREPP staff began in September 2015 and had planned to continue through June 2019 a review of programs that were reviewed under the previous criteria.¹⁰⁶

⁹⁹ "About us," The Substance Abuse and Mental Health Services Administration, accessed, May 2018, <https://www.samhsa.gov/about-us>.

¹⁰⁰ Ibid.

¹⁰¹ Ibid.

¹⁰² Ibid.

¹⁰³ Ibid.

¹⁰⁴ Ibid.

¹⁰⁵ Ibid.

¹⁰⁶ Ibid.

Review Process

NREPP identified programs for review in three ways:

- Nominations from the field - SAMHSA announced an open submission period, which generally lasted several months and allowed developers, researchers, practitioners, and other interested parties to submit programs for review.
- Environmental scans - SAMHSA and NREPP staff conducted literature searches, focus groups, public input, and interviews to identify potential interventions for review.
- Agency nomination - SAMHSA identified programs and practices addressing specific agency priorities.¹⁰⁷

Programs identified through the open submission process were prioritized for review.¹⁰⁸ Programs were pre-screened to ensure that at least one evaluation study met the minimum criteria for review and were submitted to SAMHSA for approval to move into the review process.¹⁰⁹

NREPP staff identified two certified reviewers to conduct the review.¹¹⁰ Reviewers must possess a doctoral-level degree and a strong background in and understanding of current methods for evaluating prevention and treatment interventions.¹¹¹ In addition, candidates who had direct experience providing prevention and/or treatment services were preferred.¹¹² All reviewers must have been approved by SAMHSA, and after approval, reviewers participated in a training and certification process.¹¹³

Certified reviewers were contacted to determine if they were available for a review.¹¹⁴ Assignments were made by NREPP staff, not SAMHSA, and interventions were matched with reviewers who had appropriate qualifications and the most relevant experience and content knowledge.¹¹⁵

Reviewers were required to sign a Conflict of Interest disclosure statement for each intervention they were assigned to ensure they had no professional ties or financial or other interests in the intervention that could prevent an objective review.¹¹⁶

Review packets were sent to reviewers to assess the rigor of the study and the magnitude and direction of the program's impact on eligible outcomes.¹¹⁷ Reviewers independently reviewed

¹⁰⁷ Ibid.

¹⁰⁸ Ibid.

¹⁰⁹ Ibid.

¹¹⁰ Ibid.

¹¹¹ Ibid.

¹¹² Ibid.

¹¹³ Ibid.

¹¹⁴ Ibid.

¹¹⁵ Ibid.

¹¹⁶ Ibid.

¹¹⁷ Ibid.

the studies provided and calculated ratings using the NREPP Outcome Rating Instrument.¹¹⁸ Program outcomes were reviewed on four dimensions:

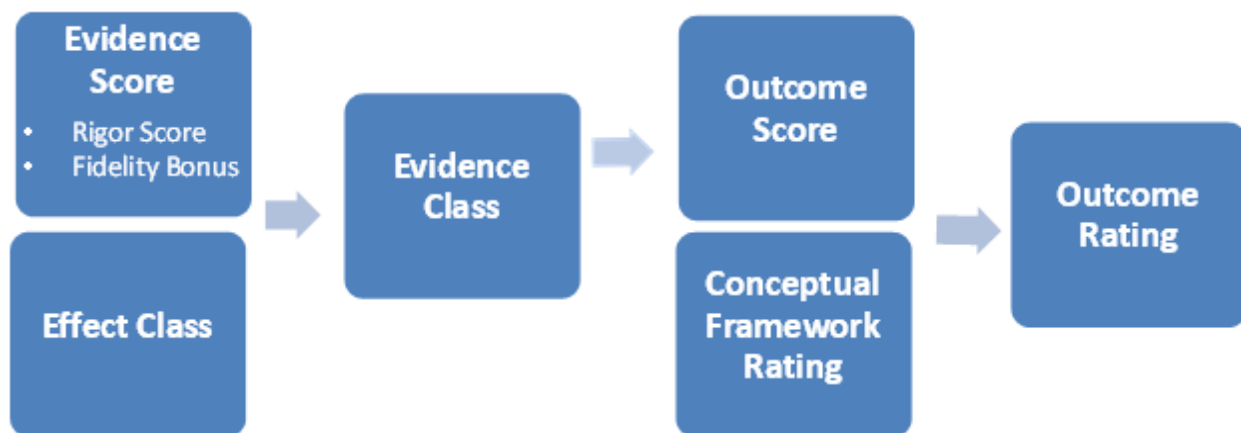
- Rigor - Rigor assessed the strength of the study methodology.
- Effect Size - An effect size is a way to measure whether a program had an impact, how big that impact was, and whether it helped or hurt the treatment group. Effect sizes are calculated when evaluation studies provide the data needed to do so.
- Program Fidelity - Reviewers examined the evaluation studies to determine if the program was delivered as intended and to the target population.
- Conceptual Framework - This dimension were concerned with how clearly the components of a program were articulated.¹¹⁹

Reviewers assigned numerical values to each dimension in the NREPP Outcome Rating Instrument (with the exception of effect size), and if the two reviewers' ratings differed by a significant margin, a consensus conference to discuss and resolve the differences was held.¹²⁰

Standard of Evidence

Although outcome ratings were based on the four dimensions, the most weight was given to the “design/assignment” element of the rigor dimension and the confidence interval of the effect size contributing to the outcome.¹²¹ Figure 3 shows the components and process used for calculating the outcome rating.

Figure 3
Components of the Final Outcome Rating



Source: https://nrepp.samhsa.gov/review_process.aspx.

¹¹⁸ Ibid.

¹¹⁹ Ibid.

¹²⁰ Ibid.

¹²¹ Ibid.

The evidence score was based on the rigor and fidelity dimensions and was rated as strong, sufficient, or inconclusive.¹²² The effect class was based on the confidence interval of the effect size:

- Favorable: Confidence interval fell completely within the favorable range (greater than 0.10);
- Possibly favorable: Confidence interval spanned both the favorable (greater than 0.10) and trivial range (from -0.25 to 0.10);
- Trivial: Confidence interval fell completely within the trivial range (from -0.25 to 0.10) or spanned the harmful and favorable range (from -0.25 to greater than 0.10);
- Possibly harmful: Confidence interval spanned both the harmful (lower than -0.25) and trivial range (from -0.25 to 0.10); and
- Harmful: Confidence interval fell completely within the harmful range (lower than -0.25).¹²³

These two dimensions were then combined to categorize programs into one of seven evidence classes, as described in Table 2.¹²⁴

Table 2 Description of Evidence Classes	
Evidence Class	Evidence Description
Class A	Highest quality evidence with confidence interval completely within the favorable range
Class B	Sufficient evidence with confidence interval completely within the favorable range
Class C	Sufficient or highest quality evidence with confidence interval spanning both the favorable and trivial ranges
Class D	Sufficient or highest quality evidence with confidence interval completely within the trivial range
Class E	Sufficient or highest quality evidence with confidence interval spanning both the harmful and trivial ranges
Class F	Sufficient or highest quality evidence with confidence interval completely within the harmful range
Class G	Limitations in the study design preclude from reporting further on the outcome

Source: Compiled by JSGC staff from https://nrepp.samhsa.gov/review_process.aspx.

¹²² Ibid.

¹²³ Ibid.

¹²⁴ Ibid.

The outcome rating was based on a combination of outcome score and conceptual framework rating.¹²⁵ The outcome score was based on the evidence class of each component measure, and the conceptual framework ratings was based on whether a program had clear goals, activities, and a theory of change.¹²⁶ Table 3 describes the outcome ratings.

Table 3 Outcome Rating	
Outcome Evidence Rating	Definition
Effective	The evaluation evidence has strong methodological rigor, and the short-term effect on this outcome is favorable. More specifically, the short-term effect favors the intervention group and the size of the effect is substantial.
Promising	The evaluation evidence has sufficient methodological rigor, and the short-term effect on this outcome is likely to be favorable. More specifically, the short-term effect favors the intervention group and the size of the effect is likely to be substantial.
Ineffective	The evaluation evidence has sufficient methodological rigor, but there is little to no short-term effect. More specifically, the short-term effect does not favor the intervention group and the size of the effect is negligible. Occasionally, the evidence indicates that there is a negative short-term effect. In these cases, the short-term effect harms the intervention group and the size of the effect is substantial.
Inconclusive	Programs may be classified as inconclusive for two reasons. First, the evaluation evidence has insufficient methodological rigor to determine the impact of the program. Second, the size of the short-term effect could not be calculated.

Source: Compiled by JSGC staff from https://nrepp.samhsa.gov/review_process.aspx.

¹²⁵ Ibid.

¹²⁶ Ibid.

The California Evidence-Based Clearinghouse for Child Welfare

Overview

The California Evidence-Based Clearinghouse for Child Welfare (CEBC) was created by the Chadwick Center for Children and Families - Rady Children's Hospital-San Diego and the Child and Adolescent Services Research Center (CASRC), and is funded by the California Department of Social Services Office of Child Abuse Prevention and is one of their targeted efforts to improve the lives of children and families served within child welfare system.¹²⁷ The CEBC is a tool for identifying, selecting, and implementing evidence-based child welfare practices that will improve child safety, increase permanency, increase family and community stability, and promote child and family well-being.¹²⁸ The CEBC provides information in simple, straightforward formats that reduce the users' need to conduct literature searches, review extensive literature, or understand and critique research methodology.¹²⁹

The CEBC is guided by three main entities:

- A statewide advisory committee comprised of state and local child welfare leaders, supporting organizations, and nationally respected authorities on child welfare;
- A national scientific panel comprised of seven core members who are nationally recognized as leaders in child welfare research and practice; and
- A national implementation science panel comprised of five core members who are nationally recognized as leaders in the field of Implementation Science and Child Welfare.¹³⁰

Review Process

The CEBC Advisory Committee is consulted annually and asked to assist in determining the topic areas that will be added to the website, and once the topic areas are chosen, a topic expert for each area is recruited with the consultation of the CEBC Scientific Panel.¹³¹ One topic expert is chosen for each topic area that appears on the CEBC website.¹³² Topic experts assist the CEBC team in identifying programs and practices that have known evidence in their area of expertise, or programs that are commonly used in California, and participate as one of the raters for programs

¹²⁷ "Overview," California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/leadership/overview/>.

¹²⁸ Ibid.

¹²⁹ Ibid.

¹³⁰ "CEBC Expert-Panels," California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/leadership/cebc-expert-panels/>.

¹³¹ "CEBC Review and Rating Process," California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/home/how-are-programs-on-the-cebc-reviewed/>.

¹³² "CEBC Topic Experts," California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/leadership/topic-experts/>.

in the topic area for which they served as an expert.¹³³ CEBC staff also conducts an extensive literature search to identify potential programs for the topic area.¹³⁴

To be reviewed, a program must have a manual or offer training, and a representative for the program must fill out a detailed questionnaire about the program.¹³⁵ CEBC staff reviews and edits the completed questionnaire and collects research studies to assemble the program outline.¹³⁶ All outcomes from research studies on a program must have been reported in a published, peer-reviewed journal to be considered by the CEBC staff during the rating process.¹³⁷

The program outlines and published, peer-reviewed research articles are sent to program raters, who typically include the topic expert and two CEBC staff members.¹³⁸ If there is a discrepancy between their ratings, the CEBC Scientific Director makes the final decision.¹³⁹ The finalized program outlines, including ratings and outcomes information, are then reviewed and posted on the CEBC website.¹⁴⁰

Standard of Evidence

There are two rating scales used on the CEBC website: the Scientific Rating Scale, and Measurement Tools Rating Scale.¹⁴¹ The Scientific Rating Scale is a rating of the strength of the research evidence supporting a practice or program.¹⁴² The Measurement Tools Rating Scale describes the CEBC ratings for tools used for screening or assessment.¹⁴³ Tables 4 and 5 describe the ratings for both of these scales.

¹³³ Ibid.

¹³⁴ “CEBC Review and Rating Process,” California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/home/how-are-programs-on-the-cebc-reviewed/>.

¹³⁵ Ibid.

¹³⁶ Ibid.

¹³⁷ Ibid.

¹³⁸ Ibid.

¹³⁹ Ibid.

¹⁴⁰ Ibid.

¹⁴¹ “Rating Scales,” California Evidence-Based Clearinghouse for Child Welfare, accessed May 2018, <http://www.cebc4cw.org/ratings/>.

¹⁴² Ibid.

¹⁴³ Ibid.

**Table 4
Scientific Rating Scale**

Rating	Rating Definition	Criteria
1	Well-Supported by Research Evidence	<ul style="list-style-type: none"> • At least 2 rigorous randomized controlled trials (RCTs) in different usual care or practice settings have found the practice to be superior to an appropriate comparison practice. • In at least one of these RCTs, the practice has shown to have a sustained effect of at least one year beyond the end of treatment, when compared to a control group.
2	Supported by Research Evidence	<ul style="list-style-type: none"> • At least one rigorous RCT in a usual care or practice setting has found the practice to be superior to an appropriate comparison practice. • In that RCT, the practice has shown to have a sustained effect of at least six months beyond the end of treatment, when compared to a control group.
3	Promising Research Evidence	<ul style="list-style-type: none"> • At least one study utilizing some form of control (e.g., untreated group, placebo group, matched wait list) has established the practice's benefit over the control, or found it to be comparable to a practice rated 3 or higher on the CEBC or superior to an appropriate comparison practice.
4	Evidence Fails to Demonstrate Effect	<ul style="list-style-type: none"> • Two or more randomized, controlled outcome studies have found that the practice has not resulted in improved outcomes, when compared to usual care. • If multiple outcome studies have been conducted, the overall weight of evidence does not support the benefit of the practice.
5	Concerning Practice	<ul style="list-style-type: none"> • If multiple outcome studies have been conducted, the overall weight of evidence suggests the intervention has a negative effect upon clients served; and/or • There is case data suggesting a risk of harm that a) was probably caused by the treatment; and b) the harm was severe and/or frequent; and/or • There is a legal or empirical basis suggesting that, compared to its likely benefits, the practice constituted a risk of harm to those receiving it.
NR	Not Able to be Rated	<ul style="list-style-type: none"> • The practice does not have any published, peer-reviewed study utilizing some form of control (e.g., untreated group, placebo group, matched wait list study) that has established the practice's benefit over the placebo, or found it to be comparable to or better than an appropriate comparison practice. • The practice does not meet criteria for any other level on the CEBC Scientific Rating Scale.

Source: Compiled by JSGC staff from <http://www.cebc4cw.org/files/OverviewOfTheCEBCScientificRatingScale.pdf>.

<p style="text-align: center;">Table 5 Measurement Tools Rating Scale</p>		
Rating	Rating Definition	Criteria
A	Psychometrics Well-Demonstrated	Two or more published, peer-reviewed studies have established the measure’s psychometrics (e.g., reliability and validity, sensitivity and specificity, etc.).
B	Psychometrics Demonstrated	One published, peer-reviewed study has established the measure’s psychometrics (e.g., reliability and validity, sensitivity and specificity, etc.).
C	Does Not Reach Acceptable Levels of Psychometrics	A preponderance of published, peer-reviewed studies has shown that the measure does not reach acceptable levels of psychometrics (e.g., reliability and validity, sensitivity and specificity, etc.).
NR	Not Able to Be Rated	Published peer-reviewed studies demonstrating the measure’s psychometrics (e.g., reliability and validity, sensitivity and specificity, etc.) are not available.

Source: Compiled by JSGC staff from <http://www.cebc4cw.org/assessment-tools/measurement-ratings/>.

The Coalition for Evidence-Based Policy

Overview

The Coalition for Evidence-Based Policy was a nonprofit, nonpartisan organization that advocated for building the body of research-proven social programs through investment in rigorous, preferably randomized, impact evaluations, and incentivized the adoption of those programs shown to produce sizable, sustained improvements in people’s lives.¹⁴⁴ The Coalition for Evidence-Based Policy wound down its operations in the spring of 2015, and the Coalition’s leadership and core elements of the group’s work have been integrated into the Laura and John Arnold Foundation.¹⁴⁵

¹⁴⁴ Jon Baron, to Board of Advisors of the Coalition for Evidence-Based Policy, April, 2015. <http://coalition4evidence.org/wp-content/uploads/2015/04/Coalition-Board-of-Advisors-Update-04-24-15.pdf>.

¹⁴⁵ Ibid.

The Promising Practices Network

Overview

The Promising Practices Network began in 1997 as a partnership between the Colorado Foundation for Families and Children, the Family and Community Trust (Missouri), the Georgia Family Connection Partnership, and the Foundation Consortium for California’s Children and Youth, four state-level organizations that help public and private organizations improve the well-being of children and families.¹⁴⁶ The founding partners created a website in 1998 with the goal of encouraging a shift towards results-oriented policy and practice by providing easier access to evidence-based information via the internet.¹⁴⁷ In 2000, the founding partners sought a fifth partner to take over website operations and content development, and ultimately chose the RAND Corporation, a nonprofit institution that helps improve policy and decision-making through research and analysis.¹⁴⁸ However, due to funding constraints, the Promising Practices Network project has concluded, and the website was archived in June 2014 and has not been updated since then.¹⁴⁹

The What Works Clearinghouse

Overview

As discussed earlier in the report, the What Works Clearinghouse (WWC) is an investment of the Institute of Education Sciences (IES) within the U.S. Department of Education that was established in 2002.¹⁵⁰ The work of the WWC is managed by a team of staff at IES and conducted under a set of contracts held by several leading firms with expertise in education, research methodology, and the dissemination of education research.¹⁵¹ The WWC is a central source of scientific evidence on education programs, products, practices, and policies.¹⁵²

The WWC reviews evidence of effectiveness of programs, policies, or practices by using a consistent and transparent set of standards, but does not rank, evaluate, or endorse interventions.¹⁵³ Hundreds of trained and certified reviewers rate whether studies meet certain standards and then summarize the results in intervention reports.¹⁵⁴ Intervention reports summarize

¹⁴⁶ “About,” Promising Practices, accessed April 2018, http://www.promisingpractices.net/about_ppn.asp.

¹⁴⁷ Ibid.

¹⁴⁸ Ibid.

¹⁴⁹ Ibid.

¹⁵⁰ “Who We Are,” The What Works Clearinghouse, accessed May 2018, <https://ies.ed.gov/ncee/wwc/WhoWeAre>.

¹⁵¹ Ibid.

¹⁵² “What We Do,” The What Works Clearinghouse, accessed May 2018, <https://ies.ed.gov/ncee/wwc/WhatWeDo>.

¹⁵³ “What Is the WWC?: A Trusted Source about What Works in Education,” *whatworks.ed.gov*, accessed May 2018, https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_info_what_061015.pdf.

¹⁵⁴ Ibid.

what the highest-quality research says about effectiveness and also include details on program components, target populations, and costs.¹⁵⁵

The WWC also develops practice guides in conjunction with an expert panel, combining the panel's expertise with the findings from rigorous research to produce specific recommendations for teaching a particular topic or addressing challenges in classrooms and schools.¹⁵⁶

When an education-related study receives media attention, the WWC issues quick reviews on the quality of the research.¹⁵⁷

Review Process

An intervention report provides a summary of findings of the highest-quality research on a program, practice, or policy in education, based on a comprehensive search of the literature.¹⁵⁸ To create an intervention report, the WWC develops a formal review protocol to identify eligible interventions, populations, outcomes, and study designs.¹⁵⁹ Then, the WWC searches systematically for all publicly available studies of the intervention, using the review protocol.¹⁶⁰ Next, the WWC reviews every eligible study against WWC standards and gives one of three ratings:

- Meets WWC Design Standards Without Reservations,
- Meets WWC Design Standards With Reservations, or
- Does Not Meet WWC Design Standards.¹⁶¹

Next, the findings from studies that meet WWC standards are combined to assess the overall effectiveness of the intervention.¹⁶² Finally, the review of the evidence is summarized in an intervention report and made available online.¹⁶³

Standard of Evidence

The WWC uses a multi-step analysis to determine whether a study meets its standard of evidence, and includes determining whether groups are randomly assigned, whether sample attrition is high or low, whether groups are similar before intervention, and whether there are confounding factors.¹⁶⁴

¹⁵⁵ "The What Works Clearinghouse," *What Works Clearinghouse (WWC)*, accessed May 2018, https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/WWC_E-Brochure_2016_022417.pdf.

¹⁵⁶ *Ibid.*

¹⁵⁷ *Ibid.*

¹⁵⁸ "What Is the WWC?: A Trusted Source about What Works in Education," *whatworks.ed.gov*, accessed May 2018, https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_info_what_061015.pdf.

¹⁵⁹ *Ibid.*

¹⁶⁰ *Ibid.*

¹⁶¹ *Ibid.*

¹⁶² *Ibid.*

¹⁶³ *Ibid.*

¹⁶⁴ *Ibid.*

The WWC also determines whether an intervention’s extent of evidence is small or medium-large, and includes this determination in the intervention report. Table 6 describes the two categories.

Table 6 Extent of Evidence	
Category	Description
Small	<ul style="list-style-type: none"> • only one study or setting, or • includes fewer than 350 students, or fewer than 14 classrooms
Medium-Large	<ul style="list-style-type: none"> • more than one study and setting, and • includes more than 350 students or 14 classroom

Source: Compiled by JSGC staff from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_info_reporting_061015.pdf.

Other Registries

A few other notable registries exist, such as the U.S. Department of Labor’s Clearinghouse for Labor Evaluation and Research, The Campbell Collaboration’s Campbell Systematic Reviews, and the University of Cambridge’s Conservation Evidence.¹⁶⁵ Like the previously-described registries, these resources provide information on various evidence-based interventions, programs, or policies designed to help decision makers implement the most effective interventions, programs, or policies. These registries employ clear standards and plain language descriptions to maximize transparency and accessibility.

¹⁶⁵ “CLEAR Clearinghouse for Labor Evaluation and Research,” United States Department of Labor, accessed May 2018, <https://clear.dol.gov>. “The Campbell Collaboration online library,” Campbell Collaboration, accessed May 2018, <https://www.campbellcollaboration.org/library.html>. “Conservation Evidence Providing Evidence to Improve Practice,” Conservation Evidence, accessed May 2018, <https://www.conservationalevidence.com/>.

PENNSYLVANIA STATE AGENCIES

As discussed, there is a significant difference between program evaluation and program measurement. In the former, a program is evaluated to determine whether it achieves the desired outcomes for which it was intended. In the latter, a program is measured to gather data and information about the resources it consumes in providing its services and to determine how many people are beneficiaries of its services, for example.

Commonwealth agencies spend a great deal of time and resources in exercises of program measurement. The annual Governor's Budget proposal contains hundreds of pages of Commonwealth agency program measures that are submitted by the various state entities. Vast amounts of data are collected and analyzed to provide decision makers with the information they need to manage the Commonwealth's enormous obligations to its residents. EBDM is accomplished in its truest sense when scientific research methods are applied to public policy decision making. Such endeavors are costly, resource intensive, and not quickly concluded.

For example, the Pennsylvania Board of Probation and Parole (PBPP) is an agency that does indeed conduct EBDM processes. The PBPP contracts with research firms that provide analyses of dozens of programs and applies statistical methods to help create algorithms that assist officers in working with parolees to achieve best outcomes.

The process is successful, but to maintain that level of success the PBPP invests heavily in its contractual relationships with the research firms, its own IT capability, its own data gathering, and most importantly, in the continuing education of its staff and officers.

Pennsylvania's executive branch agencies are continuously engaged in gathering tremendous amounts of data and information. Each agency, given its unique responsibilities, collects and interprets information from its constituent populations, whether from licensed drivers, state-chartered banks, recipients of medical assistance, or the Commonwealth's deer herd.

Because budgeting for Commonwealth functions goes hand in hand with carrying out those functions, measuring the amount of money spent is one of the foundational means by which program parameters may be set. Thus, each of the overarching means of oversight focuses on funding resources.

Each year, Pennsylvania's so-called "budget season" opens with the governor presenting his proposed budget for the upcoming fiscal year. The Governor's Executive Budget presents nearly 1,000 pages of state program descriptions, measures, past funding, requested funding, and levels of funding expected in ensuing years. This document is a compendium of program measures and is widely used as a reference.

Program Measures in the Governor's Executive Budget. The Commonwealth currently follows a "program budgeting" process, which frames funding decisions in terms of the objectives that may be achieved through activities underwritten by state revenues. A distinguishing feature of the program budgeting approach is the inclusion of Program Measures in the Governor's Executive Budget document. Program Measures provide information on progress toward stated objectives that decision makers may consider when allocating resources through the annual budget process. The Governor's Executive Budget for the 2018-19 fiscal year includes approximately 600 Program Measures that provide data on inputs, activities, outputs, efficiency, and outcomes associated with Commonwealth programs. **EBDM components included in the Executive Budget: program assessment, budget development, and outcome monitoring.**

Report on State Performance. Since 2008, the Governor's Budget Office has published an annual Report on State Performance. The purpose of the report is to provide Pennsylvania's citizens and public officials with information on the results achieved through the investment of public resources. The most recent editions of the reports are organized into seven goal areas: Education, Economic Development, Health and Human Services, Environment, Consumer Protection, and Government Efficiency. Within these goal areas, the report presents measurable objectives, strategies for achieving priority results, and performance measures for tracking progress toward targets. **EBDM components included in the Report on State Performance: program assessment, budget development, and outcome monitoring.**

OpenDataPA. On April 18, 2016, Governor Wolf signed an executive order instructing the Office of Administration (OA) to create and maintain an enterprise open data portal for the Commonwealth. Pennsylvania's first open data portal launched later in 2016 as OpenDataPA. "Open data" refers to data in a format that can be understood by a computer and used freely by anyone. OpenDataPA functions as a central repository for sharing the Commonwealth's data with the public, entrepreneurs, civic developers, researchers, and policy makers. The portal allows users to measure government performance, engage with state government, explore economic opportunities, and develop innovative solutions to public policy challenges. To date, more than 70 datasets have been made publicly available on OpenDataPA. **EBDM components included in the OpenDataPA portal: program assessment, budget development, implementation oversight, outcome monitoring, and targeted evaluation.**

Enterprise Data Sharing Memorandum of Understanding (MOU). To facilitate responsible data sharing among agencies under the Governor's jurisdiction, OA established an Enterprise Data Sharing MOU in 2017. Prior to creation of the Enterprise Data Sharing MOU, when agencies wanted to share data with one another, they would develop an MOU for each dataset. Over the years, this resulted in hundreds of agreements with inconsistent terms, conditions, and other provisions. The Enterprise Data Sharing MOU eliminates the time and effort to create individual MOUs and brings consistency to data sharing agreements. The new process streamlines interagency data exchanges and demonstrates the Wolf administration's commitment to the value of data sharing, while maintaining the integrity, security, and quality of Commonwealth data. **EBDM components included in the MOU: program assessment, budget development, implementation oversight, outcome monitoring, and targeted evaluation.**

Performance-Based Budgeting. On October 30, 2017, Governor Wolf signed the Performance-Based Budgeting and Tax Credit Efficiency Act into law. The Act instructs the Independent Fiscal Office (IFO) to conduct reviews of performance information for all Commonwealth agencies and for all tax credits at least once every five years. The IFO is formally required to develop performance-based budget and tax credit plans that recommend approaches to improving the effectiveness and efficiency of Commonwealth agencies' operations and programs. The IFO's recommendations are to be taken under consideration by the Governor and the General Assembly during the annual budget development process. **EBDM components included in Performance-Based Budgeting: budget development, implementation oversight, outcome monitoring, and targeted evaluation.**

Office of Performance through Excellence (OPE). On February 1, 2018, Governor Wolf signed an executive order creating the Office of Performance through Excellence (OPE). OPE will combine the previous efforts of the Governor's Office of Transformation, Innovation, Management, and Efficiency (GO-TIME) with the Wolf administration's active use of performance management and lean management.¹⁶⁶ The office's purpose is to engage the Commonwealth's workforce to modernize operations, maximize efficiency, and provide the highest quality services to the people of Pennsylvania. OPE embodies the Wolf administration's commitment to making policy decisions based on evidence and to pursuing continuous improvement in state government operations. **EBDM components included in the OPE: program assessment, budget development, implementation oversight, outcome monitoring, and targeted evaluation.**

Commission staff contacted each agency's legislative liaison office to request information about their EBDM processes. The agencies under the governor's jurisdiction coordinated their responses through the governor's policy office. The information that was provided is presented below.

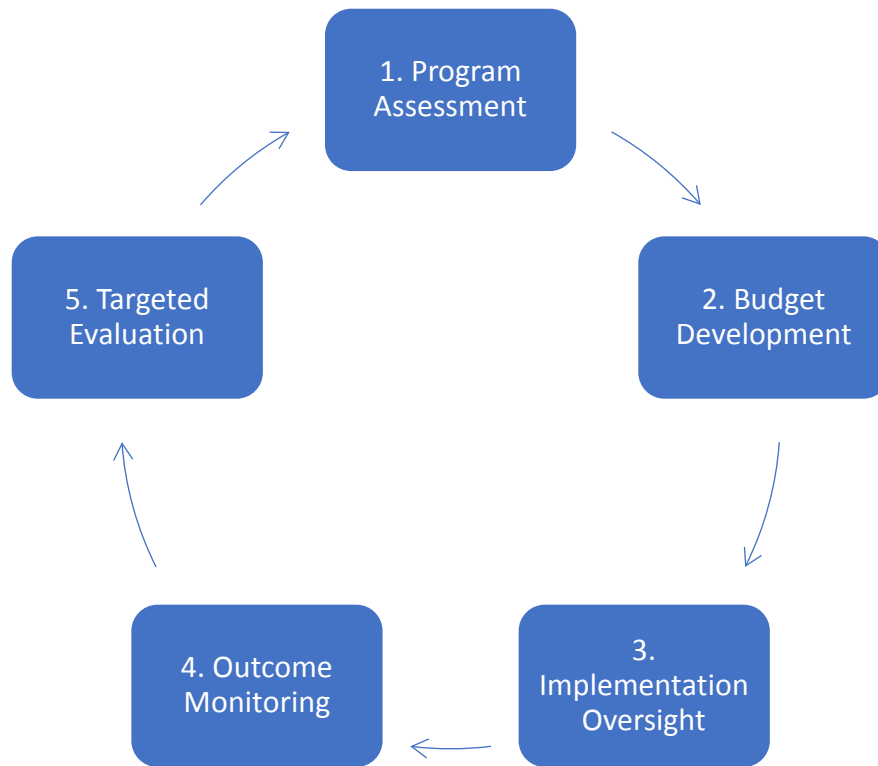
Pew Foundation

The Pew Foundation created the "Results First Initiative" as an approach to help state governments engage in EBDM. Recognizing that empirical research into public policies requires funding and time that often constrain states from engaging in it, Pew Charitable Trusts and the MacArthur Foundation are collaborating to build a clearinghouse of states' data and information in order to leverage the collective knowledge in support of evidence-based policy making. Assuming that the Commonwealth is able to gather and analyze outcomes data and evaluations, there are concrete steps that can be taken as listed by Pew.

The Pew-MacArthur Results First Initiative uses a framework consisting of five components shown in Figure 4. Each component is associated with governmental functions in an iterative EBDM process.

¹⁶⁶ Lean management, "seeks to eliminate any waste of time, effort or money by identifying each step in a business process and then revising or cutting out steps that do not create value." TechTarget, "lean management," <https://searchcio.techtarget.com/definition/lean-management>, accessed July 5, 2018.

Figure 4
Results First Framework



Each of the five steps involves data collection and interpretation. The first three components, program assessment, budget development, and implementation oversight are (or should be) core functions of government program administrators regardless of whether the entity is engaged in EBDM.

Program Assessment

Assessment begins with taking inventory of the agency’s programs and collecting information to establish each program’s baseline performance measures. Researchers and decision makers use the information developed in this first stage of the cycle to identify which programs are worthy of further investigation and whether selected programs are achieving desired outcomes. That is, they ascertain why programs may be effective, how the resources might be spent more effectively, or what innovative interventions may provide improved outcomes.

Pew recommends that researchers and decision makers rank the assessed programs based on the strength of the evidence that has been gathered for each. In the Pew “Results First” system, those programs that exhibit effectiveness by way of the most rigorous testing methods are categorized as “evidence-based.” A second tier category, labeled “promising programs,” includes

“those that have been evaluated and shown effective but through a less rigorous research design.”¹⁶⁷

Return on Investment.

One of the components in Pew’s system of program assessment directs decision makers to apply cost-benefit analysis to programs, “to rank programs by their potential return on investment...”¹⁶⁸ Public sector return on investment calculations include both financial data and non-financial data that have been monetized. Various techniques exist for monetizing non-financial outcomes such as one would find in quality of life measures.¹⁶⁹

Budget Development

Resource management, particularly funding, is one of the primary drivers of the move toward EBDM. Program assessments and cost-benefit analyses from the first stage of Pew’s Results First framework, information analysis, program goals, and outcomes are the focus of decision makers’ efforts to direct funding to the most effective programs while diverting it from those programs that are not meeting expectations or serving their purposes. The task of budget development is best accomplished when the parties responsible for each program are in open communication with one another and share necessary information.

Further, decision makers can leverage their budgetary powers to improve programs and foster innovation by providing incentives to program staff for meeting or exceeding expected outcomes. As Pew recommends, “[w]hen practicable, contracts and grants should include performance goals that encourage organizations to provide evidence-based programs and to implement those services as designed.”¹⁷⁰

Implementation Oversight and Outcome Monitoring

One area in particular in which state agencies apply a lot of their staff resources is in the oversight of program implementation. Program monitoring involves regular contact between headquarters staff and program staff, whether within agency or reaching out to contractors and grantees. Arguably, while all involved recognize the importance of monitoring and oversight, the most common complaint from service providers in the field is about the burdensome time and effort spent on complying with oversight requirements. The paperwork burden, in many cases, (and especially in human services), threatens to overcome program staff’s work with its program’s intended objectives and goals. With thin personnel ranks and increasing expectations of positive outcomes, program providers are fearful they will find themselves spending more time on paperwork than in contact with clients in drug and alcohol treatment settings, for example.

¹⁶⁷ Gary VanLandingham, et al. “Evidence-Based Policymaking: A Guide for Effective Government,” *Pew MacArthur*, 5. <http://www.pewtrusts.org/~media/assets/2014/11/evidencebasedpolicymakingaguideforeffectivegovernment.pdf>.

¹⁶⁸ *Ibid.* 7

¹⁶⁹ Washington State Institute for Public Policy. Benefit-cost technical documentation. Olympia, WA: Author. (December 2017), <http://www.wsipp.wa.gov/BenefitCost>. Accessed May 21, 2018.

¹⁷⁰ Gary VanLandingham, et al. “Evidence-Based Policymaking: A Guide for Effective Government,” *Pew MacArthur*, 9. <http://www.pewtrusts.org/~media/assets/2014/11/evidencebasedpolicymakingaguideforeffectivegovernment.pdf>.

Nonetheless, EBDM relies on data gathered from field operations. Whether measured as contact hours with SUD clients, life expectancies of patients, recidivism rates of juvenile offenders, or the number of motor vehicles registered, decision makers cannot make informed decisions—indeed cannot follow any system of EBDM—without having evidence in hand. Further, without oversight of implementation of programs to ascertain whether those programs are being efficiently delivered, whether they show fidelity to model, or whether they achieve the expected outcomes, evidence-based decisions cannot be made.

Targeted Evaluation

The fifth, or end, stage of the process is also the first, or initial stage of the process. Targeted evaluations occur when state agencies “[c]onduct rigorous evaluations of new and untested programs to ensure that they warrant continued funding.”¹⁷¹ Because the scientific method is iterative, meaning that the overall process is evaluated and modified with respect to the goals in parallel with the research itself, targeted evaluation of a program loops back to program assessment, budget development, etc.

The key objective in beginning the cycle is that the agencies need to first develop a priority list for programs that they intend to study. Agencies’ work in performance measurement, i.e. the tasks that result in information such as that provided in the governor’s annual budget proposal, can be used to help agency executives create priority lists for their agencies’ performance evaluations. As mentioned previously, performance evaluations are expensive undertakings. Pew recognizes that state government research offices:

[C]an be a critical resource in conducting independent program evaluations, but historically, much of their work has focused on assessing compliance and management issues rather than outcomes. Legislators can work with these offices as they set their research agendas to identify opportunities to dedicate a larger portion of their resources to determining whether programs are achieving desired results.¹⁷²

The Pew-MacArthur Foundation report, *How States Engage in Evidence-Based Policymaking*, divides states into four categories: Trailing, Modest, Established, and Leading.¹⁷³ This categorization is based on their implementation of EBDM. Four policy areas were examined in determining the rankings. Those policy areas were:

- Behavioral Health
 - Programs to improve mental health
 - Decrease substance abuse

¹⁷¹ Gary VanLandingham, et al. “Evidence-Based Policymaking: A Guide for Effective Government,” Pew MacArthur, 5. <http://www.pewtrusts.org/~media/assets/2014/11/evidencebasedpolicymakingguideforeffectivegovernment.pdf>.

¹⁷² Ibid.

¹⁷³ Elizabeth Davies, et al, ““How States Engage in Evidence-Based Policymaking,”” The Pew-MacArthur Foundation, January 2017, http://www.pewtrusts.org/~media/assets/2017/01/how_states_engage_in_evidence_based_policymaking.pdf.

- Child Welfare
 - Reduce incidents of children mistreatment
- Criminal Justice
 - Reduce recidivism of convicted offenders
- Juvenile Justice
 - Reduce recidivism of adjudicated youth

The report ranked Pennsylvania’s use of EBDM as Modest. This score, however, is an aggregate of the four policy areas examined according to Pew’s criteria, and is not meant to imply that Pennsylvania’s programs in Behavioral Health, Child Welfare, Criminal Justice, and Juvenile Justice are ineffective. In the area of Criminal Justice, for example, the report explains that states “often had guidelines defining evidence-based criminal justice programs as those with a ‘proven’ ability to reduce recidivism, without specifying what standard of evidence is needed to meet these criteria.”¹⁷⁴

Although the report does go on to warn that “[s]uch ambiguous definitions can become problematic for states trying to identify and fund evidence-based programs consistently.”¹⁷⁵ The lack of consistency and accuracy in defining evidence-based policies was noted by Bret Bucklen, director of research and statistics at the Pennsylvania Department of Corrections:

“[The] term is thrown around quite loosely within our system, [and] in some cases the evidence behind so-called evidence-based programs is quite weak. Our challenge is to help staff understand what constitutes strong evidence, such as randomized controlled trials.”¹⁷⁶

Further, a significant consideration is that Pennsylvania’s primary responsibility is to create and deliver programs and services to most effectively address the Commonwealth’s needs. A commitment to meeting the Pew-MacArthur parameters is, quite understandably, secondary. Nonetheless, it can be helpful for decision makers to sharpen and specify the term EBDM to achieve desired results. The creation of strong data infrastructure and the analytical and technical capacity to take advantage of it are essential to provide policymakers with the tools to utilize EBDM. At the level of program outputs, such investments will help evaluate performance measures and enable program fidelity. None of the advancements are possible, however, without the state’s dedication of resources and funding.

¹⁷⁴ Ibid. 27.

¹⁷⁵ Ibid. 27.

¹⁷⁶ Ibid. 27.

http://www.pewtrusts.org/~media/assets/2017/01/how_states_engage_in_evidence_based_policymaking.pdfhttp://www.pewtrusts.org/~media/assets/2017/01/how_states_engage_in_evidence_based_policymaking.pdf.

Executive Branch Agencies

Office of Administration (OA)

OA is responsible for data sharing across the Commonwealth. OA is using automated IT capabilities to assist in its project. IT capabilities assist the implementation of monitoring through tracking activity, status, and project health. IT assists in solving unplanned service interruptions, reduction of its occurrence and management change. OA implemented new metrics for the current fiscal year across both HR and IT services. The new metrics include:

For HR

- Cost and efficiency
- Customer Satisfaction
- Hiring process effectiveness
- Quality of candidates
- Recruitment effectiveness
- Classification effectiveness
- Employee engagement
- Workforce planning
- Talent development effectiveness
- Safety effectiveness
- Service utilization
- Employee relations

For IT

- Projects at risk
- Cost efficiency
- Value and Customer Satisfaction
- Project timeliness
- Projects completed within budget
- Application availability
- Infrastructure availability
- Retirement eligibility risk
- Retirement trends
- Cyber incident resolution rate
- Security awareness
- Privileged users
- Security compliance

Examples from OA potentially span all five components of EBDM, as OA is responsible for data sharing across the Commonwealth, both internally and externally (see reference to the Enterprise Data Sharing MOU in the letter dated March 2, 2018). OA is using automated IT capabilities to assist with project portfolio management. These capabilities assist with implementation oversight through activity tracking, status, and project health. Similarly, OA is using a tool to automate IT service management processes. These capabilities assist with implementation oversight and outcome monitoring through IT asset and configuration tracking, along with resolving unplanned service interruptions, reducing reoccurrences, and change management. Connections to budget development spring from IT asset tracking.¹⁷⁷

¹⁷⁷ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Office of the Budget (Budget Office)

The Budget Office prepares a biannual strategic plan that includes goals, strategies, measurable objectives, performance targets, and metrics. Metrics in the current plan include:

- Financial Collections
- Cost reduction
- Cost avoidance
- Customer satisfaction
- Employee turnover
- Employee satisfaction
- Value-added or risk based services offered
- Process efficiency and effectiveness

Office of Inspector General (OIG)

The Office of Inspector General (OIG) tracks and reports several outcome metrics for its programs. These reports are used as part of the annual budget development process. OIG employees are obligated to complete an online study course that leads to certification as a fraud examiner. Also, staff is encouraged to attend ongoing training to stay up-to-date with best practices. Moreover, OIG participates as a member of Association of Inspectors General (AIG), a non-profit organization that promotes excellence in the inspector general profession.

Program Assessment and Implementation Oversight

Staff in the Bureau of Special Investigations have certifications through the Association of Certified Fraud Examiners (ACFE). They complete an online study course and certification test that allows them to become certified fraud examiners. The OIG also participates as a member of the AIG. The AIG is a non-profit, membership organization for agencies and professionals in the inspector general community. Their mission is to promote:

Excellence in the inspector general community by establishing and encouraging adherence to quality standards, sponsoring professional development and networking opportunities, certifying individuals in IG-specific disciplines, supporting offices in governmental and external relations, and inspiring governmental entities to embrace the inspector general model as an effective tool in the fight to combat waste, fraud and abuse.

OIG staff can attend trainings and conferences for both the ACFE and the AIG to network and stay up-to-date with best practices being used throughout the country.

Budget Development and Outcome Monitoring

OIG tracks and reports several outcome metrics for their programs, which are considered as part of the annual budget development process. The metrics include measures of collections of overpaid benefits and cost savings to the Commonwealth resulting from investigative activities and prosecutions.¹⁷⁸

¹⁷⁸ Letter from Colby Clabaugh, Executive Budget Specialist, Office of the Budget to Joint State Government Commission, March 15, 2018.

Department of Agriculture (PDA)

PDA uses a process of EBDM to turn challenges into opportunities and to confront threats to Pennsylvania agriculture. Two very serious threats facing Pennsylvania's agriculture economy are the recent arrival of the spotted lanternfly (SLF) and the continuing threat of highly pathogenic avian influenza (HPAI).

SLF is an invasive species that can cause catastrophic damage to crops, the ecosystem, and overall quality of life. PDA responded by organizing a board of state and federal experts who are charged to remove or mitigate the threat as much as possible. The board meets once a month to discuss data, information, setbacks, and progress against the spread of SLF. The board's decisions are based on evidentiary findings; methods in the field are flexible and change depending on what works. The plans are based on scientific research largely from Penn State University, where work has been ongoing for over three years. One newly formed program is focused on SLF containment, which focuses on developing trap trees instead of conducting tree removals, as previously had been attempted.¹⁷⁹ The program, based on years of experimental trials, is cost effective and more efficient than previous efforts.

The HPAI program focuses on protecting both animal and human health. Multiple organizations monitor the program, including the Bureau of Animal Health and Diagnostic Services within PDA. Data provided from multiple universities determine program methods used to effectively eradicate HPAI and other diseases. As a result of the unpredictability of the animal diseases and infections, program budgets are based on experience and evidence generated from both in- and out-of-state resources.

Program Assessment

SLF containment. A team composed of staff from PDA's Bureau of Plant Industry (including the State Entomologist and the State Plant Regulatory Official) and USDA's Animal and Plant Health Inspection Service (APHIS) meet at least monthly to review and revise, where necessary, a plan to contain the SLF. The plan is based on research conducted by researchers from Penn State University and other higher-education institutions, as well as a USDA researcher.

HPAI monitoring. Protecting both animal and human health is a critical role of PDA's Bureau of Animal Health and Diagnostic Services. Programs within this bureau rely heavily on science and research to determine actions.

¹⁷⁹ Trap trees are trees the trunks of which are wrapped with sticky tape-like bands that trap SLF as they crawl up and down the trees' trunks. PennState Extension, "Spotted Lanternfly Management: Placing Sticky Bands on Trees," <https://extension.psu.edu/spotted-lanternfly-management-placing-sticky-bands-on-trees>, accessed July 5, 2018.

Budget Development

SLF containment. Funding for SLF containment efforts is based on research findings and trials over the three years that Pennsylvania has been working to contain this pest. For example, research and trials have determined that removing the Ailanthus tree from every location where the SLF is identified is not effective in containing the SLF. In addition, tree removal is time consuming and costly. Therefore, the focus in year four of the program will be on developing trap trees and not so heavily focused on tree removal.

HPAI. Because of the unpredictability of animal disease and infection, the budget for this bureau is based on experience and practice and relies heavily on evidence and data from both in-state and out-of-state resources.

Implementation Oversight

SLF containment. This program is heavily monitored as the SLF has significantly impacted the agriculture industry, specifically crops of grapes, hops, and apples. The pest creates quality of life issues, as people in the core of the quarantine zone have found themselves at times to be unable to leave their homes or go outdoors due to the SLF's heavy presence.

HPAI. If this program is not effectively delivered, serious impacts on human and animal health could occur. The HPAI program is monitored by the Bureau of Animal Health and Diagnostic Services and a committee with membership from the veterinary medical field, producers, advocacy organizations, and governmental entities.

Outcome Monitoring

SLF containment. The program's frequent meetings provide an opportunity to report outcome data and to determine if the program should be modified in any way or if additional funding is required.

HPAI. Phone calls are held monthly with a Poultry Health committee to discuss the status of avian health in Pennsylvania, the United States, and around the world. The effectiveness of the biosecurity plans, testing, reporting, and on-going surveillance is discussed to determine if additional or corrective action should occur.

Targeted Evaluation

SLF containment. Ongoing research provides the program team with information on whether current program efforts are likely to continue to be effective, or new methods should be implemented.

HPAI. Pennsylvania relies on a laboratory network consisting of the Department of Agriculture, the University of Pennsylvania, and Penn State University. Data from these three labs provide PDA with tools to determine the effectiveness of this program.¹⁸⁰

Department of Banking and Securities (DoBS)

DoBS regulates financial services and works to ensure that consumers and businesses are well-informed about the banking and securities marketplaces. As a regulatory body, DoBS monitors the Commonwealth's financial institutions to protect consumers and investors based on dozens of measures and indicators. The department itself strives to meet many performance targets. Notable performance targets are those listed below.

Program Assessment and Implementation Oversight

DoBS is accredited by three independent organizations: Conference of State Bank Supervisors, American Association of Residential Mortgage Regulators, and National Association of State Credit Union Supervisors. As such, the department undergoes a comprehensive and thorough review and evaluation of its program areas as well as its departmental administrative functions.

Budget Development and Outcome Monitoring

Current DoBS performance targets and measures are as follows:

- Respond to consumer complaints in less than 15 days;
- Examine a minimum of 20 percent of all Securities Investment Adviser registrants per year, which results in all registrants being examined every five years;
- Examine a minimum of 20 percent of all non-depository licensees per year, which results in all licensees being examined every five years;
- Ensure that 90 percent of non-depository examiners have the highest level of certification;
- Ensure that 90 percent of depository examiners have the highest level of certification; and
- Issue the final report of depository examinations within 30 days of when the final draft is submitted to the Harrisburg Office.

The DoBS performance measures listed above are included in the department's annual budget request and the Governor's Executive Budget.

¹⁸⁰ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Targeted Evaluation

DoBS is a regulator, and as such the department's programs (and duties) are established by national standards and are being constantly evaluated or redesigned in response to the ever-changing financial industry and markets.¹⁸¹

Department of Community and Economic Development (DCED)

DCED assess its programs by measuring data collected from each of nearly 100 programs, each with different tasks and responsibilities. DCED reviews its guidelines and enforces compliance based on those guidelines. DCED trains employees in lean management so as to ensure that the department's programs operate with effectiveness and efficiency. DCED empowered its Performance Monitoring Division to oversee results from its business programs. The intent is to measure the program's performance and hold accountable individual grant and loan recipients. DCED requires full repayment of grants from any company that moves operations outside of Pennsylvania.

Program Assessment

Program managers review and assess each program. Achievements are measured through data collection specific to intended outcomes for each program objective. There is currently no assessment to compare outcomes of programs against each other. DCED has close to 100 programs and each of these serves a specific need for the department's customers. Program assistance often comes from a combination of DCED's programs. The programs are not mutually exclusive.

Implementation Oversight

On an annual basis, DCED reviews its guidelines to make sure programs are implemented and delivered according to the intended design.

At the time of this report, DCED undertook a comprehensive internal control assessment in response to Management Directive 325.12, which directed Commonwealth agencies under the governor's jurisdiction to comply with the federal government's Green Book to ensure that each agency designs, implements, and operates internal controls to achieve its objectives related to operations, reporting, and compliance.

Finally, DCED had begun training employees in lean management, with the intent of administering the department's programs in an efficient and outcome-driven manner.

Outcome Monitoring

DCED has a dedicated Performance Monitoring Division that monitors the results of many of its business financing programs. The intent is to measure program performance, but also to hold individual grant and loan recipients accountable for the taxpayer investment they receive. DCED,

¹⁸¹ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

through its Performance Monitoring Division, already collects millions of dollars annually from companies who fail to meet their previously agreed-upon commitments. These policies have been further strengthened with changes that have recently gone into effect. For example, if a company receives a grant from the Commonwealth and subsequently moves operations out of Pennsylvania, DCED will require full repayment of the grant, as well as a 10-percent penalty.

Targeted Evaluation

This is currently done on a case-by-case basis. DCED does not have a systematic approach to evaluate new and untested programs, but the department expressed willingness to explore ideas on how this can best be accomplished.¹⁸²

Department of Conservation and Natural Resources (DCNR)

DCNR implements EBDM by collecting data and evaluating their performance. For each plan, data are gathered and analyzed. The data collected by DCNR are open to private citizens, businesses, and government agencies.

DCNR deals with various programs such as the number of state park campsites rented each year, number of LEED certified buildings, Statewide Comprehensive Outdoor Recreation Plan data (SCORP) and Pennsylvania Natural Diversity Inventory. SCORP is one of several programs required to update plans every five years in order to receive federal Land and Water Conservation Fund funding. The data gathered for each plan includes a statewide survey of recreational priorities and economic data on recreational activities that generate revenue. The information is used to promote hiking, trail use, and walkable communities.

Implementation Oversight and Outcome Monitoring

Water well data. The Pennsylvania Ground Water Information System (PaGWIS) database houses data on water well locations, construction, yields, and the geologic units penetrated by each well. The data are openly available and are used by businesses, government agencies, academics, and private citizens.

Oil and gas well data. The Exploration and Development Well Information Network (EDWIN) database houses data on oil and gas well locations, construction, production, and geologic units penetrated by each well. The data are obtained through the DEP permitting process and transferred to the Bureau of Topographic and Geologic Survey (BTGS) for inclusion in the database. The data are openly available and are used by businesses, government agencies, academics, and private citizens.

Geologic data. BTGS geologists collect a variety of geologic data in the course of the department's work. These data are released to the public in the form of maps, reports, GIS datasets, and searchable databases. The data are openly available and are used by businesses, government agencies, academics, and private citizens.

¹⁸² Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Pennsylvania Natural Diversity Inventory (PNDI). In 2016, DCNR responded to public comments, many from industry, to upgrade its PNDI process for reviewing and identifying potential impacts to threatened and endangered species that are part of DEP permit reviews. DCNR created and put online the PA Conservation Explorer Tool, a GIS database that allows any user, at no cost, to input potential project outlines and determine what, if any, impacts it will have to listed species or species of concern. This has enabled users to shift their projects or redesign them to avoid impacts, a win for project developers and for the environment. It has also increased transparency. For those users who complete the process of applying for a permit with a PNDI receipt, DCNR instituted a \$40 fee that helps pay for updating the data. By all accounts, users have been pleased with the new tool and project review times have been reduced substantially. Data on plant and animal species come from DCNR, the Game Commission, the Fish and Boat Commission, U.S. Fish and Wildlife, academic and nonprofit partners, and are updated periodically.

DCNR staff collect and use a wide variety of data streams to inform their day-to-day programs and policies, and to operate more efficiently and cost-effectively.

Number of state park campsites rented each year. Over years of tracking this performance measure, DCNR found that some camp sites are not used year-round. Metrics showed that DCNR's full-service hook-up sites were being reserved a year in advance, and that demand for these sites had been increasing every year. DCNR has been converting hundreds of underused tent sites to full-service hook-up sites each year. Each converted site generates an additional \$3,356 per site, per year, which yields a 343 percent increase in revenue. Every 300 sites converted generates \$1 million in additional annual revenue.

Number of LEED-certified buildings in DCNR's state parks and state forests. In the early 2000s, DCNR began designing LEED-certified buildings when constructing or renovating its park and forest buildings for energy efficiency and to model sustainable building elements. At the time of this report, DCNR had 15 completed LEED-certified buildings, with a goal of completing 20 buildings by 2020. DCNR reviewed this metric to make sure LEED was appropriate in all cases and—with feedback from district staff, maintenance staff and department engineers—created a high-performance building policy in 2017 to gauge when it pays to use LEED or another process. With help from the Department of General Services, (DGS) through the Guaranteed Energy Savings Act (GESA) audit program, DCNR is converting one-third of all buildings to achieve zero emissions and lower energy and water use.

Statewide Comprehensive Outdoor Recreation Plan (SCORP) data. SCORP is updated every five years and is a requirement to continue receiving federal Land and Water Conservation Fund monies. Data gathered for each plan include a statewide survey of recreational priorities for Pennsylvania citizens, economic data on how recreational activities generate revenue for nearby communities and businesses, and other topics. DCNR's current plan, which covers 2014 to 2018, includes data on health benefits related to outdoor recreation. The department uses this information to promote more hiking, trail use, and walkable communities, as well as in grant making.¹⁸³

¹⁸³ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Department of Corrections (DOC)

The DOC has demonstrated its commitment to EBDM over years of using empirical methods to fulfill its mission to “reduce criminal behavior by providing individualized treatment and education to inmates, resulting in successful community reintegration through accountability and positive change.”¹⁸⁴ The DOC works closely with PBPP to develop and implement programs based on results derived from multiple research reviews from outside sources and its own internal research. DOC and PBPP initiated over 100 randomized controlled trials (RCTs). Both the DOC and PBPP utilize agreed-upon key performance metrics as part of the annual budget process. Each important metric has a baseline and five years to achieve its goals. The DOC ensures its efficiency by process evaluations and program monitoring. The Bureau of Planning, Research, and Statistics provides an online clearinghouse of research reports that present available evidence on relevant topics in Corrections and Criminal Justice.

Program Assessment

The department’s Bureau of Planning, Research, and Statistics performs a number of functions that are critical to EBDM in the department:

- coordinates departmental research initiatives
- prepares various planning and research reports based on correctional databases
- implements standards, guidelines and procedures for state research and data analysis activities
- identifies and defines correctional planning and research problems
- prepares research models appropriate to the problem
- develops instrumentation and procedures for data measurement
- collects and analyzes data
- submits recommendations for departmental research and evaluation priorities

Importantly, these are the types of activities that all large government entities should follow in applying EBDM.

Budget Development

As cited above, DOC utilizes agreed-upon key performance metrics in determining both its own budget and funding for PBPP as part of the budget process. Each key metric has a baseline and a five-year target.

Implementation Oversight

Recent implementation oversight included reviews of an array of DOC programs across the spectrum of the department’s obligation to public safety and inmate needs. The department

¹⁸⁴ “About us,” Department of Corrections, accessed May 2018, <http://www.cor.pa.gov/About%20Us/Pages/CONTACT%20US%20-%20About%20Us.aspx>.

utilizes monthly information provided through the Justice Reinvestment Initiative (JRI), which is a partnership between The Pew Charitable Trusts and the U.S. Department of Justice, Bureau of Justice Assistance. DOC also uses information from the monthly Joint Performance Measures (JPM) online dashboard. The JRI digital dashboard was developed jointly by PCCD, DOC, PBPP, PaSDC and the Council of State Governments (CSG) Justice Center. It provides key monthly and annual performance measures that track the outcomes associated with the JRI in Pennsylvania, including inmate population, admissions and releases, reinvestment of savings, crime trends and recidivism rates.¹⁸⁵

Oversight and evaluation activities included:

- Process evaluation of the DOC's Transitional Housing Units (THUs);
- Evaluation of Community Corrections, and in-prison treatment programs' fidelity of model, using the Correctional Program Assessment Inventory (CPAI) and the Correctional Program Checklist (CPC);
- Support of George Mason University's process evaluation of the department's DOC's Therapeutic Communities (TCs); and
- Monitoring compliance and implementation fidelity with the Disability Rights Network (DRN) and Department of Justice (DOJ) lawsuit settlement agreements on DOC's mental health practices, using the Central Office Special Needs Psychiatric Review Team (COSNPRT) report.

Outcome Monitoring

DOC is legislatively mandated to document outcomes in the biannual Recidivism Risk Reduction Incentive (RRRI) program performance report, the State Intermediate Punishment (SIP) program performance report, and the Quehanna Boot Camp program performance report. Also, the department implemented performance-based contracting scorecard with all contracted Community Corrections facilities.

Targeted Evaluation

DOC has initiated over 100 rapid, staff-proposed randomized controlled trials (RCTs) of ideas for:

- Reducing in-prison violence;
- Reducing use of solitary confinement; and
- Improving staff wellness.

BetaGov, a research group from New York University, is providing technical assistance in the development, implementation, and reporting of these rapid RCTs. Recently the project was

¹⁸⁵ PA Justice System Joint Performance Measures website, <https://jridashboardpa.net>, accessed July 11, 2018.

extended to another goal of developing RCT trials for improving community-based corrections, parole supervision, and reentry. At least two dozen trial ideas have currently been proposed in this area.

An RCT evaluation of Vivitrol, an extended-release antagonist drug for treating opioid and alcohol abuse was conducted in partnership with the University of Pennsylvania and Drexel University. Other RCT evaluations included:

- whether reentry is best facilitated by returning parolees to their home community or to a new community;
- a study referred to as the “Relocation Study,” conducted in conjunction with University of Maryland; and
- an RCT evaluation of a program to provide Pell grants to fund college courses for inmates, which is being conducted in conjunction with University of Maryland and Drexel University.

The department developed a priority agenda for rigorously evaluating all of DOC’s and PBPP’s programs and practices. This list currently includes 94 programs and practices. Four evaluations will be conducted each year. Currently the first program evaluation is being completed, which is an evaluation of in-prison Therapeutic Communities’ impact on recidivism. A cost-benefit analysis is included in each evaluation.¹⁸⁶ The department partnered with Drexel University in an evaluation of the SIP/HOPE disciplinary program within the SIP program in Community Corrections Centers. Other recent research included evaluations of:

- the impact of family visitations on inmate recidivism;
- the use of solitary confinement, conducted with Harvard University and Columbia University;
- recidivism outcomes for various options currently available for sanctioning technical parole violators (TPVs); and
- State Intermediate Punishment and County Intermediate Punishment.

Department of Drug and Alcohol Programs (DDAP)

DDAP is the Commonwealth’s Cabinet level agency that provides treatment and rehabilitation services to Pennsylvania residents who suffer from substance use disorders and gambling problems. Among its myriad responsibilities, the department is one of several state and federal agencies that monitor, inspect, license, and fund service providers. DDAP coordinates the continued improvement of existing and the development of new treatment modalities and programs.

¹⁸⁶ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Most of the funding administered by DDAP is received by the Single County Authorities (SCAs), which coordinate local provision of treatment and prevention services for persons suffering from substance use disorders and gambling problems. DDAP enforces guidelines to ensure programs are following the intended design designed by the Bureau of Administration and Program Support.

The performance evaluation by DDAP is limited because of the current implementation of a new data system for treatment and prevention activities. However, DDAP still tracks part of the SCA performance using the legacy system. The legacy system tracks results of needs assessments rather than effectiveness of interventions. The collected data and reports are sent to the National Outcome Measurement System (NOMS) at SAMHSA to monitor the accomplishments and measure the effectiveness of the services.

Budget Development

Assessment of programming, particularly as it relates to the funding of the SCAs, may affect funding levels as part of the SCA allocation process. The budget at large is less affected by such adjustments in funding since the bulk of the funds received by DDAP are issued to the SCAs to provide substance abuse and gambling problem prevention and treatment services, and that overall amount is designated at a specific level within the budget. However, that does not preclude adjustments among SCAs based on expenditure capability or factors like compliance or performance.

Implementation Oversight

Mechanisms in place to ensure that programs are effectively delivering services according to the intended design include the work completed by the County Program Oversight (CPO) Section in the Bureau of Administration and Program Support, as well as the monitoring of activities by the Division of Prevention and Intervention in the Bureau of Treatment, Prevention and Intervention through the Performance Based Prevention System (PBPS).

The annual monitoring process completed by CPO consists of a combination of desk reviews for materials and reports submitted by the SCAs, as well as an annual on-site review during the period beginning in May and extending through October. In addition to the monitoring of the SCAs carried out by the department, CPO provides the SCAs with an instrument for monitoring their contracted providers. As part of the monitoring conducted by CPO, a sample of the monitoring completed by the SCAs is reviewed to ensure compliance with this requirement, as well as to assess the SCAs in addressing any provider deficiencies.

Generally, the monitoring conducted by the SCAs exceeds the minimum requirements issued by the department in content and frequency, and includes both a programmatic and fiscal review. Additionally, the Division of Prevention uses PBPS, a process-oriented data system, to track the SCAs' needs assessment, planning, and service implementation through their contracted prevention providers. This system allows staff to suggest or enforce adjustments, as needed, to the approved plan, and to contracting and service delivery.

Outcome Monitoring

DDAP is currently implementing a new data system for treatment and prevention activities. Until this system is fully implemented, the department's ability to monitor outcomes for some of the measures identified in the grant agreements with the SCAs, specifically related to treatment, will be challenging. Given that the legacy system applicable to prevention is still operational, certain measures are still tracked. These measures provide information on the application of recommended programming that aligns with the results of the needs assessment, rather than on the effectiveness of interventions. However, pre- and post-test information is available through the system that provides some information on the success of the educational efforts used for prevention. Additionally, data are collected and reported relative to both treatment and prevention in reporting performance indicators and accomplishments under the National Outcome Measures System. These measures provide an indication of success (or failure) in various areas relative to the effectiveness of services.

Targeted Evaluation

Most funds disbursed by the department are received by the SCA network, an established system of county-based or county-related entities responsible for the assessment, planning, and delivery of services for their local geographic areas. However, it is not uncommon to evaluate new programs and initiatives for effectiveness and efficiency, and to determine the merits of continuing the programs, as well as identifying barriers that limit the potential of such initiatives. Additionally, SCAs have an obligation to evaluate contracted providers to ensure effective and efficient utilization of public funds.¹⁸⁷

Department of Education (PDE)

PDE uses EBDM processes to fulfill its mission of providing high quality education to Pennsylvanians.

Program Assessment

PDE is committed to promoting equity, innovation, and transparency, and believes that decisions and initiatives that impact students, educators, and communities should be rooted in rigorous research and evidence, and aligned to Pennsylvania's academic standards.

Pennsylvania's Consolidated State Plan under the federal Every Student Succeeds Act (ESSA), approved by the U.S. Department of Education in January 2018, identifies investing in evidence-based strategies as a top priority. ESSA places significant emphasis on evidence-based interventions and strategies, and requires that states, LEAs, and relevant partners identify activities, strategies, or interventions that demonstrate statistically significant effects on improving academic and other relevant outcomes.

¹⁸⁷ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Implementation Oversight and Outcome Monitoring

In addition to its focus on improving resources and fair funding for public K-12 schools, the department has also worked diligently to improve the measures of school performance in preparing students for postsecondary success, shared through a new dashboard-style report card. The Future Ready PA Index is expected to launch in fall 2018, and will include nearly two dozen research-based indicators of school performance, including student growth, progress toward English language proficiency, chronic absenteeism rates, career readiness indicators, attainment of industry credentials, access to well-rounded and advanced course offerings, and postsecondary outcomes, among others.

Targeted Evaluation

PDE has also worked in recent months to provide relevant, timely analysis on priority issues, like computer science education and postsecondary enrollment and attainment, through publicly available data visualizations.¹⁸⁸

Department of Environmental Protection (DEP)

DEP is working to streamline permit applications, clarify permit regulatory requirements, and progress in permit backlog. DEP is seeking to improve its oil and gas resource management by restructuring its permit programs, improve workload, and review consistency. DEP's attempts to improve its work efficiency led to establishment of internal process evaluating to fulfill staff vacancies. Also, it led to the creation of performance data to analyze budget development process.

DEP is developing a data dashboard to allow for significant improvement in program oversight. Moreover, it identifies key staff needs and supports developing, planning management, and compliance to guidelines. The board meets semiannually to ensure consistent implementation of the program and its efficiency. Through the implementation of various programs, DEP manages to evaluate and enforce efficiency in programs.

DEP successfully solved several issues relating to permits and improved the efficiency of its program. Also, it managed to modernize its complaint tracking system. DEP initiated a newly formed plan called Pittsburgh 100, which aims to reduce inefficiency and backlogs. Another new program is SWDO, which reduced review time by 220 days and cut the backlog in half.

Program Assessment

DEP is issuing new and revised technical guidance documents (TGDs), improved permit application forms, and instructions to clarify regulatory requirements to permit applicants and facilitate the submission of adequate permit applications to the department. Examples include the Comprehensive Environmental Assessment TGD, Draft Water Quality Certification Process for Interstate Gas Transmission Pipeline Projects TGD, and Aquatic Resource Functional Assessment Protocols. These TGDs help the regulated community submit complete and accurate permit

¹⁸⁸ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

applications the first time, thus streamlining the review process and allowing DEP to allocate resources efficiently.

DEP has made tremendous progress in reducing the overall permit backlog across programs. As of November 2017, the queue was reduced by 1,544 permits; permit applications overdue for review were reduced by 2,164; and permit applications overdue for completeness review have been reduced by 2,454.

The Office of Oil and Gas Management is restructuring its permitting programs across the three districts, which review and issue well and surface permits. The purpose of this restructuring effort is to equalize the permit review workload and improve the permit review consistency among the three districts by assigning one statewide program manager to oversee all surface permit reviews statewide and one program manager to oversee all well permits statewide. Once fully implemented, this initiative will improve the permit review process and enhance protection of resources while reducing permit review backlogs.

Budget Development

The department uses performance data in the budget development process. For example, DEP has observed that a decline in staffing for the Safe Drinking Water program since 2009 has been accompanied by declining performance across several key indicators, including the number of sanitary surveys (full inspections) conducted at public water systems, the percentage of community water systems meeting health-based drinking water standards, and the number of unaddressed violations. DEP maintains that the data support a direct relationship between performance and available resources for the Safe Drinking Water program.

The department established an internal process to evaluate and prioritize filling of staff vacancies as opportunities arise. The department has been managing its complement as responsibly as possible. However, General Fund cuts over the last decade have been detrimental to the department even as it shoulders significant new responsibilities for oversight. Prioritizing vacancy fillings is critical to fulfilling the department's mission to protect health, safety, and the environment.

Implementation Oversight

DEP is in the process of developing a data dashboard that will allow reviewers and managers at all levels to have significantly greater oversight of the programs for which they are responsible. The data dashboard, which will be operational by mid-2018, will give managers a clearer picture of backlogs and current permit timeframes, and help identify staffing needs and determine vacancy allocations among different offices.

Central office staff provides technical and administrative support to the core duties of implementation, planning, management, compliance, and permitting in the regional offices. The support work involves developing, interpreting, and providing constructive input on regulations, policies, procedures, and technical guidance for programs, as well as developing, assessing, and delivering technical training programs for regional office and local government staff.

Section Chiefs and Program Managers meet semi-annually to ensure consistent implementation and efficiency in their respective programs.

The Office of Program Integration (OPI) is responsible for improving efficiency in the implementation of environmental protection programs, coordinating efforts that impact several programs within the agency, and ensuring consistency in the implementation of programs statewide. Additionally, under the purview of OPI, Standard Operating Procedures (SOPs) were developed in a manner to provide for consistent permit review regardless of which region/district an application is submitted. SOPs are organized sequentially by activities that will be completed.

Outcome Monitoring

The department audited all the permits in the centralized electronic system, known as eFACTS, and discovered an unreasonable number of permits that were not tracked all the way through to completion. This can create confusion about where in the process permit applications stand. The department received in excess of 40,000 permits in 2017, the majority of which were entered into eFACTS. Since the summer 2017, DEP has corrected and made current its internal permit tracking system.

DEP is continuously evaluating environmental reporting procedures for ways to increase efficiency in processes without impacting environmental protection. Many of these improvements have occurred within the Oil and Gas program, but are continuing to expand throughout the agency in areas related to permitting, inspections, payment collection, bidding, and grants.

DEP tracks complaints that are received across all programs through an internal complaint tracking system. In an effort to capture information that is solely related to complaints from individuals who suspect that their drinking water supply has been affected by oil and gas related activities, DEP modernized the department's complaint tracking system and added a component to track oil and gas related water supply complaints.

Since 2015, DEP has been working to modernize permitting activities through an electronic permitting initiative. Currently, underground mining permits and storage tank renewal authorizations may be applied for online. This has helped to reduce incomplete application submittals—which take up valuable permitting time—and has made permit tracking easier. The Oil and Gas program estimates that an electronic well permit application will help reduce well permit review times by 30 percent.

DEP will use a phased approach to add several permit applications across programs to the ePermitting system by the end of 2018, thereby facilitating additional wait- time reductions across programs. Through four phased steps, DEP will add the following programs to the ePermitting system by the end of 2018:

- Office of Oil and Gas Management
- Bureau of Air Quality
- Bureau of Radiation Protection
- Information Technology
- Bureau of Waterways Engineering and Wetlands
- Office of Active and Abandoned Mining Operations
- Division of Storage Tanks
- Bureau of Clean Water

Targeted Evaluation

DEP conducts rigorous evaluations of all programs, old and new, to ensure that these programs are efficient and effective, while meeting required regulatory obligation.

In May 2017, DEP's Southwest District Office (SWDO) implemented an innovative new permit review approach referred to as the Pittsburgh 100. The goal of this program is to shrink review timelines of the Erosion and Sediment Control General Permit (ESCGP), a key development permit, to under 100 days while reducing backlogs for well permits. DEP is achieving this by:

- Auditing the performance of SWDO, while bringing new leadership to the office;
- Rotating pending permits to regions with capacity;
- Reallocating positions within the Office of Oil and Gas Management in the SWDO;
- Targeting new hires in key permit review areas; and
- Providing permit reviewers with additional training to standardize permit review processes.

As a result of these steps, DEP is achieving the ambitious goal of the Pittsburgh 100. Since implementation of these key steps, the SWDO has reduced its permit review timeline for the ESCGP by over 220 days, while cutting the backlog in half. Additionally, a backlog of well permit applications exceeding DEP's review timelines has been reduced from 289 in August 2017 to 36 as of March 2018. The department will continue to implement the successful lessons of the Pittsburgh 100 to improve timeframes and reduce backlogs.¹⁸⁹

¹⁸⁹ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Department of Health (DOH)

DOH is responsible for three broad public health functions: assessment, policy development, and assurance. Further, DOH is responsible for ten essential public health services, including:

- Monitoring health status to identify and solve community health problems
- Diagnose and investigate health problems and health hazards
- Inform, educate, and empower people about health issues
- Mobilize community partnerships and action to identify health problems
- Enforce laws and regulations that protect health and ensure safety
- Link people to needed personal health services
- Assure competent public and personal health care workforce
- Evaluate effectiveness, accessibility, and quality of personal and population-based health services
- Research for new insights and innovative solution to health problems

DOH programs are evaluated on cost effectiveness and performance. Federal funding requires DOH to submit program evaluation and performance reports annually. As a means to improve its outreach to the public and professionals, DOH adopted Enterprise Data Dissemination Informatics Exchange (EDDIE). The EDDIE program is a web-based interactive health tool through which users create data tables and share with others.

The department publishes regular State Health Improvement Plans (SHIP). The current plan identifies three health priorities: 1) obesity, physical inactivity and nutrition; 2) primary care and preventive services; 3) mental health and substance abuse. Under Act 191 of 2014, DOH implemented Pennsylvania's Prescription Drug Monitoring Program (PDMP).¹⁹⁰ Results appear to show a 12.6 percent decrease in opioid prescribing in the PDMP's first year of full implementation. Last year, DOH adopted Tuberculosis Screening and Treatment's pilot program to provide better service to Pennsylvania citizens and save money through cost-effective plans.

Department of Human Services (DHS)

DHS is responsible for providing social services and supports to vulnerable residents of Pennsylvania. DHS conducts performance measurements of all of its programs. For example, Healthcare Effectiveness Data & Information (HEDIS) and PeopleStat analyze data and provide important performance measure information about DHS programs. HEDIS is used by 90 percent of America's health plans because it measures performance and provides data that can be used to identify problems. PeopleStat is a data driven process that reviews outcomes and assists DHS in ensuring programs meeting their requirements.

¹⁹⁰ Act of Oct. 27, 2014, (P.L.2911, No. 191), "Achieving Better Care by Monitoring All Prescriptions Program (ABC-MAP) Act."

DHS pays Managed Care Organizations (MCOs) to efficiently provide the public with quality services. DHS includes pay-for-performance standards as part of funding MCOs. Also, DHS enforces financial penalties on MCO plans that perform poorly. DHS signed intergovernmental agreements with University of Pittsburgh to provide DHS with empirical evidence on programs' efficiency and methods to improve performance.

The Keystone STARS program seeks to improve the quality of early childhood learning. Pennsylvania's Quality Rating and Improvement System is used to assess, improve, and communicate the level of quality in early and school-age care and education programs that are part of Keystone STARS.

The Community Health Choices program assists Medicaid recipients who have physical disabilities, the objective being to coordinate health care services to provide quality services to its recipients. To improve performance, Community Health Choices established an evaluation plan to better understand the organization of the program and allow for further improvements.

Budget Development

Many DHS programs are federal entitlement programs, which poses obstacles to reallocation of funding despite what EBDM might indicate. As such, the department takes other steps to ensure these programs are effective and that funds are allocated responsibly. There are several ways the department achieves this, including:

- In DHS's Medicaid managed care contracts, the department includes pay-for-performance standards as part of the payment process. The managed care organizations get paid for meeting specific performance standards in terms of their members' health outcomes. In addition, managed care plans that perform poorly in specific member health outcomes may be required to pay financial penalties. By focusing on health outcomes, DHS ensures that dollars are being spent for quality services and not just for delivering services. DHS is focusing more on these types of arrangements by requiring the managed care plans to have a specific amount of their spending be on value-based purchasing within its provider networks. This includes provider pay-for-performance, patient-centered medical homes, and bundled payments.
- Payments to training providers participating in the Temporary Aid for Needy Families (TANF) program have been based, in part, on a person obtaining employment. The department continues to review this payment process and revise it accordingly to ensure it places the right emphasis on both job placement and a person's ability to maintain employment.
- For child care, the department has focused a lot of attention on standards for quality early learning by establishing the Keystone STARS program. This program is Pennsylvania's Quality Rating and Improvement System (QRIS). A QRIS is a continuous quality improvement systemic approach to assess, improve, and communicate the level of quality in early and school-age care and education programs. The department has placed specific emphasis on expanding the number of children receiving subsidized child care into facilities with higher STARS rating.

- The department has recently proposed providing \$4 million to expand evidence-based home visiting programs focusing on families with children that are struggling with opioid addiction. These evidence-based programs focus on increasing parental skills, reducing child abuse, enhancing child development, and improving child health outcomes.

Implementation Oversight

The department regularly reviews data on its programs, especially by the entities that are responsible for the day-to-day administration of the programs. The department administers many of its programs through competitively procured contracts or through county governments. As such, it is important for DHS to review the data at a vendor or county level. In doing so, the department can see which entities may not be performing well and require interventions or corrective actions. This data review also helps the department identify which entities may be performing well, and thus help identify promising practices that can be used by others. Some examples of this include:

- For DHS's physical and behavioral health Medicaid managed care plans, staff regularly reviews the Healthcare Effectiveness Data and Information (HEDIS) measures. HEDIS is a tool used by more than 90 percent of America's health plans to measure performance on important dimensions of care and service. Altogether, HEDIS consists of 94 measures across seven domains of care. Using these data, the department can measure how well Pennsylvania is doing compared to other states, and also compare among the different managed care plans. HEDIS data allow DHS to identify areas of poor performance, and when necessary, request a plan of correction from a managed care plan.
- In Pennsylvania, the Child Welfare System is operated by the local county children and youth agencies. The department provides state and federal funding for the county children and youth agencies to work with children and their families. As such, the department is regularly reviewing data from each county, including child placements, reunifications with families, and other measures for child safety and well-being. The department uses the federally required dataset, which each county must use to report to the Commonwealth, to determine areas of potential improvement. When a county is struggling in a certain metric related to child welfare, DHS will intervene with that county to discuss what steps can be used to improve performance.

Outcome Monitoring

DHS has a regular, data-driven process to review outcomes and data to ensure that department programs are meeting their established objectives. This process, called PeopleStat, is based upon five goals for the department. Each goal area has assigned metrics and targets that are reviewed with department executive staff at meetings that occur every other month. The regular review of data allows DHS executive staff to understand the performance of its programs and be proactive in implementing interventions for areas where performance may not be meeting the identified targets.

Targeted Evaluation

The department regularly reviews programs to determine their continued effectiveness. Some examples of this include:

- The department has an intergovernmental agreement with the University of Pittsburgh's Medicaid Research Center to provide rigorous, science-based evaluations of Pennsylvania's Medicaid program. These evaluations allow the department to understand utilization, outcomes, and spending patterns in the Medicaid program and potential interventions to improve the delivery of care to Medicaid recipients.
- The department also recently evaluated the effectiveness of its Keystone STARS program for child-care settings. This evaluation (referred to as "re-visioning") aims to ensure that the STARS program continues to identify and promote quality child care programs that improve the early childhood learning.
- Recently, the department began to roll out managed long-term care for Medicaid recipients who are aging or have a physical disability and need long-term care services and supports. This program, called Community Health Choices, aims to coordinate health care services and encourage services to be provided in a person's home. With the roll out of the program, the department established an evaluation plan to better understand how this program is operating and to allow for any necessary future adjustments.¹⁹¹

Department of Labor and Industry (L&I)

L&I is responsible for economic development, working conditions, and benefits to unemployed individuals. L&I programs are located in the Bureaus of Workforce Partnership and Operations (BWPO) and Workforce Development Administration (BWDA), and the Office of Vocational Rehabilitation (OVR). BWPO administers employment services under Title 111 of the Workforce Innovation Act, and provides access to employment and related services for its customers. All of the BWPO programs are all federally funded.

BWDA maintains an inventory of funded programs. Moreover, it provides training and technical assistance to each of its programs to improve the efficiency and performance. BWDA conducts monthly conference calls and on-site visits to ensure the quality of the service provided and efficiency. The Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC) prepare RSA-113, which provides quarterly information on individuals with disabilities to the Rehabilitation Services Administration (RSA). The data are used to track and analyze applications for VR services and to identify people with significant disabilities to determine eligibility. Thus, the budget of OVR and HGAC depends on market trends, student needs, and ancillary supportive services.

¹⁹¹ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

The Workers' Compensation and Insurance Deputate ensures that all business process, and system automation revisions are meeting functional standards and expectation. It publishes weekly reports that identify key input and output data for program areas.

Unemployment Compensation Service Centers. The Unemployment Compensation Service Centers use data analysis to review the effectiveness of the delivery of unemployment compensation benefits, adhering to a philosophy of "Plan, Do, Check, Act," and to adjust how they performs as needed.

Bureau of Disability Determination (BDD). The BDD utilizes software and hardware owned by the federal Social Security Administration to ensure efficiency in providing data to federal government and other entities.

Program Assessment

Bureau of Workforce Partnership and Operations (BWPO). The BWPO administers the Employment Services under Title III of the Workforce Innovation and Opportunity Act (WIOA) on behalf of the U.S. Department of Labor's Employment and Training Administration (USDOL ETA). BWPO also administers the Trade Act program, Rapid Response services, Foreign Labor Certification and Migrant Seasonal Farmworker programs, Workforce Opportunity Tax Credit program, National Dislocated Worker Grants, and the Jobs for Veterans State Grants (JVSG). These programs are all federally funded and monitored by USDOL. The programs are, for the most part, offered through the 58 PA CareerLink offices throughout the Commonwealth.

Bureau of Workforce Development Administration (BWDA). The BWDA maintains an inventory of funded programs.

Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC). HGAC assesses its fifteen educational program offerings and ancillary supportive services at the end of every semester.

Budget Development

Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC). Data steer the budget process at the HGAC in the following ways:

- Closing and opening educational programs based on market trends and employment/placement outcomes;
- Shifting staffing patterns to support needs of the students; and
- Closing and opening ancillary support services to maximize service use and student involvement.

Implementation Oversight

Workers' Compensation and Insurance Deputate. Workers' Compensation conducts weekly program status update meetings and regularly scheduled key external stakeholder conference calls to ensure all business process, OIT, and system automation revisions are achieving functionality standards and expectations. The Workers' Compensation programs also generate a weekly report identifying key input and output data for program areas. These indicators provide a weekly assessment of key factors within the business process.

Unemployment Compensation Service Centers. Unemployment Insurance utilizes hourly, daily, weekly, and monthly datasets to systematically review the effectiveness of the delivery of unemployment compensation benefits. These data points range from hourly call rates to federally mandated measures of performance. Unemployment Insurance also seeks internal and external customer feedback regarding whether the program is meeting their needs. If their needs are not met, adjustments are made. Additionally, Unemployment Insurance uses a "Plan, Do, Check, Act" philosophy to determine when adjustments are needed based on receiving feedback that is inconsistent with expectations. There are daily, weekly, monthly, quarterly, and annual reports produced to aid in these analyses.

Bureau of Disability Determination (BDD). BDD utilizes software and hardware owned or directed by the federal Social Security Administration (SSA) to ensure effective delivery of all programs and services. BDD ensures effective implementation by maintaining contact with SSA and components impacted by the SSA programs that BDD implements, including L&I, U. S. Congressional offices, Pennsylvania legislative offices, and the legal services community.

Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC). The OVR Systems and Evaluation Division prepare the RSA-113 Report, which provides for the quarterly collection and reporting of information on individuals with disabilities receiving services through OVR. The report is submitted quarterly to the Rehabilitation Services Administration (RSA) of the federal Department of Education. Data from the RSA-113 are used to track trends of persons applying for VR services, determinations of eligibility made by OVR, identification of persons with significant disabilities, individualized plan for employment (IPE) development, service implementation, and program outcomes. These trends provide key program measurements and inform a general assessment of the state- federal VR program and its effectiveness.

- *Pre-employment Transition Service (PETS).* Work-based learning activities contracts are reviewed using a monitoring tool that assesses performance relative to contract specifications. Act 26 data, reported by counties, are reviewed to determine what is working and what needs improvement.
- Other contracts, agreements, and Letters of Understanding are monitored using a monitoring tool, and when available, a Report Card that is generated out of the Commonwealth Workforce Development System (CWDS) case management system.
- *Bureau of Workforce Partnership and Operations (BWPO).* BWPO reviews USDOL requirements for all programs and ensures that all field offices follow established procedures. The bureau's central office staff also provides technical assistance to field

staff when entering data on the CWDS for all programs. Field offices hold regular staff meetings to review processes and provide feedback to the central office for possible improvements. Monthly reports are available from CWDS for review.

- *Bureau of Workforce Development Administration (BWDA)*. BWDA provides training and technical assistance to grantees to improve program outcomes. BWDA holds monthly conference calls with grantees based on the project to ensure goals and objectives are reviewed, and where necessary, to assist in developing a corrective action plan to meet outcomes. BWDA also collects pre- and post- evaluation forms on demonstration projects, as well as quarterly progress reports for projects awarded under WIOA statewide activities funds and re-employment funds. The bureau conducts annual on-site monitoring for OMB Uniform Guidance Compliance and conducts WIOA compliance monitoring either via desk review or on-site.

Outcome Monitoring

Bureau of Disability Determination (BDD). SSA continually monitors outcome data to ensure effective management of the disability program, and provides ongoing outcome reporting to BDD and other federal governmental units and entities.

Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC). The OVR Systems and Evaluation Division prepare the RSA-911 Report, which collects and reports specific data elements for each OVR customer to the RSA. Data collected and reported include credentials and skills obtained while participating in the VR program and employment outcome information such as employer, occupation, hourly wage, and hours worked per week. OVR uses the data collected through the RSA-911 for myriad program evaluation and reporting purposes, for both internal and external stakeholders. RSA uses the data to describe the performance of the VR and Supported Employment programs in their Annual Report to Congress and the President as required by the Rehabilitation Act of 1973, as well as to assess the performance of the VR program through the calculation of evaluation standards and performance indicators codified in the Rehabilitation Act and consistent with the common performance accountability measures established under Title I of WIOA for the core programs of the workforce development system.

HGAC. Outcomes regarding employment placement are monitored, updated, and reported quarterly to the leadership team at HGAC, which uses the data to shape and guide strategic planning, educational programming, ancillary support programming, and engagement. Outcome data from consumer and stakeholder evaluations are used to monitor effectiveness and results of 18 departments and 15 educational programs.

Bureau of Workforce Development Administration (BWDA). For discretionary grants, L&I reviews and negotiates outcomes in the grantee's project proposals at time of award and then reviews quarterly and final progress reports on performance and financial outputs. Under WIOA, outcomes may be determined by local workforce boards.

Targeted Evaluation

Bureau of Disability Determination (BDD). SSA evaluates program quality and effectiveness on an ongoing basis, and makes all funding determinations, including ongoing project and initiative funding.

Office of Vocational Rehabilitation (OVR) and Hiram G. Andrews Center (HGAC). A Vocational Rehabilitation Specialist is responsible for conducting and compiling all evaluations of educational and support programming at HGAC. Data from all evaluations are used to further expand effective programming, or close and shift resources to successful or needed programming at HGAC.¹⁹²

Department of Revenue (DOR)

DOR's mission is to efficiently administer Pennsylvania's tax laws and other revenue programs. DOR's Strategic Plan Goal is to accurately and timely analyze revenue receipts and forecasts. DOR evaluates its progress and standards with agency metrics. The department engages in analytics and evaluations that support decision making through "revenue generation" projections, improved business processes, and technology modernization and innovations. To ensure the efficiency and verify the correct implementation of its programs, DOR uses Independent Verification and Validation (IV&V) to provide checks and balances on outside contractors.

Program Assessment

Strategic Plan Goal 4 is to provide accurate and timely revenue analysis. Achieving this goal requires the department to develop the necessary analyses, models, and projections to enable the Governor's Office, the General Assembly, and other interested parties to make informed decisions about revenue matters. The department is constantly evaluating and reviewing the requirements of its core mission. DOR evaluates its progress against projections and standards with agency metrics. Many of the metrics are "best practices" used by many state revenue agencies across the country.

Budget Development

Strategic Plan Goals 3, 7, and 8 have DOR utilizing analytics and reporting to yield information about existing policies. DOR incorporates EBDM in various programs and policies. As an example, the largest department expenditure is personnel. When determining what area to invest resources in, the department uses an evaluation model that incorporates "revenue generation" as a component to managing its personnel complement.

Implementation Oversight

As outlined in the DOR Strategic Plan Goal 3, the department will continually improve business processes and technology. Modernizing technology, innovation, and replacing legacy

¹⁹² Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

systems requires constant oversight to ensure that funds are disbursed properly and to prevent waste, fraud, and abuse. In order to verify that implementation maintains fidelity to design and intent, the department uses Independent Verification and Validation (IV&V) services on major IT projects, and empowers an internal modernization program leadership team to provide checks and balances on outside contractors.

Outcome Monitoring

The department measures outcomes with agency metrics. Examples of DOR's use of outcome data to inform decision making are as follows:

- *Call center wait times.* The department's Taxpayer Service and Information Center (TSIC) is a call center designed to be a point of entry for taxpayers to receive general and taxpayer-specific information. In fiscal year 2014-15, the average telephone wait time was approximately 11 minutes, 12 seconds. This was one of the highest average wait times experienced by taxpayers in years. To combat the increase, the department deployed several interventions to improve internal metrics, and also to provide better customer service. One intervention used was the deployment of the Automated Dialer system, also known as Robo Calls, to relay messages to Property Tax Rent Rebate (PTRR) recipients. Over 291,000 automated calls were made notifying claimants that their applications had been received. In addition, over 319,000 of the 404,000 (79 percent) total claimants whose applications were scheduled to be paid on time received an automated call. This intervention, in conjunction with several others, enabled TSIC to lower its call center wait time by several minutes.
- *Motor fuels electronic filing rates.* The Department of Revenue's Bureau of Motor and Alternative Fuel Taxes (MAFT) instituted an electronic filing requirement in late 2016 for all Pennsylvania registered fuel distributors. This requirement mandated that all registered fuel distributors must file electronically, and then subsequently, MAFT would reduce the bond requirement. The electronic reporting resulted in an improvement in reporting efficiency and return accuracy. By lowering the bond requirement, the fuel distributors saved money through lower operating costs. At the outset of the announcement, approximately 238, or 43 percent, of the 549 required monthly liquid fuels and fuel taxpayers were filing electronically via the department's online E-TIDES tool. After the intervention went into effect, the number of electronic filers increased to 99 percent.

Bulk mailing project. Reducing both the cost and the administrative burden associated with bulk mailing of paper forms is a priority and one of the goals of the electronic filing initiative. In 2016, the department implemented a public-facing online bulk tax forms ordering system for use by legislators, libraries, post offices, CPAs, VITA sites, and county Registers of Wills and Recorder of Deeds offices. The intent was to streamline the bulk tax forms ordering process, expedite order fulfillment, and provide order tracking. This was a notable operational accomplishment over the process in previous years. The department mailed 1,951,790 forms in 2016 at a cost of more than \$350,000, which includes temporary clerical support staff costs. Unfortunately, less than 100,000 of these forms were used. The remaining forms that were not returned for processing were unused and presumably wasted. The low usage of these paper forms

calls for an evaluation of the cost and effort used to provide this service. As a courtesy service, the department has traditionally provided these forms to stakeholders. The results of this project highlight that this service may no longer be relevant. The next steps in this project will be to evaluate stakeholders' needs and develop an outreach and education campaign that outlines DOR's electronic filing channels. As taxpayer behavior changes, the department must be prepared to evolve and meet the needs of its customers.¹⁹³

Department of State (DOS)

DOS administers a wide array of functions related to professional licensing, corporate registries, elections, campaign finance, and lobbying disclosure. The Bureau of Commissions, Elections, and Legislation compares voter registration applications by application source (online, motor voter, agency-based, paper). It measures the cost effectiveness of the programs based on the following: average processing time, percentage of applications approved, and percentage of applications rejected. It also monitors phone call statistics on a weekly basis to ensure efficiency and resource allocation.

Professional Health Monitoring Program's (PHMP) primary mission is to protect the public. PHMP reports outcomes on as-requested basis. The report analyzes data and addresses violations. PHMP measures effectiveness by successful completions of programs. The Bureau of Professional and Occupational Affairs (BPOA) conducts evaluations of similar programs to narrow scopes of work and training. BPOA also conducts thorough and coordinated audit processes.

Budget Development

Bureau of Commissions, Elections, and Legislation. The bureau compares voter registration applications by application source (online, motor voter, agency-based, paper, voter registration drive) to measure cost-effectiveness based on the following criteria: average processing time, percentage of applications approved, percentage of applications rejected, and percentage of applications pending. The bureau also compares notary applications by application source (online, bulk online via third-party provider, and paper) to measure the cost-effectiveness of its online and bulk notary solutions, based on the following criteria: average processing time, percentage of applications approved, and percentage of applications rejected.

Bureau of Enforcement and Investigations (BEI). BEI monitors its workload and turnaround time to support budget development efforts and ensure there are sufficient resources to enforce the laws, rules, and regulations of the 29 boards and commissions that BEI is tasked with providing oversight to.

¹⁹³ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Implementation Oversight

Bureau of Commissions, Elections, and Legislation. The bureau monitors the following on at least a weekly basis to determine program area effectiveness and resource allocation:

- Phone call statistics, including the number of calls handled by program area, the percentage of calls handled within the “answer time” standard, the percentage of calls abandoned, and the disposition or nature of each call;
- The number of notary public applications processed and the average processing time; and
- The number of certifications, apostilles processed, and the average processing time.

Bureau of Corporations and Charitable Organizations (BCCO). BCCO ensures effective delivery of its customers’ filings to meet the expectation of keeping processing times at or below three business days. Overtime is also implemented during peak filing periods or unusually large filing days. BCCO incorporated expedited filings with instructions that the filer must submit their filing in person. There are no exceptions to this filing policy. In 2017, the bureau processed 10,699 expedited filings (61 one-hour filings, 502 three-hour filings, and 10,136 same-day expedited filings).

Bureau of Enforcement and Investigation (BEI). BEI monitors turnaround time to ensure that businesses are inspected in a timely manner and on a routine basis. Moreover, BEI strives to conduct new business inspections as quickly as possible to guarantee that the business can open without any unnecessary down time for the licensee.

Bureau of Professional and Occupational Affairs (BPOA). BPOA ensures that the licensed professional maintains the highest standards of practice throughout the active license period. BPOA also conducts a thorough and coordinated audit process, ensuring that the professional licensee maintains a high standard of education while keeping up with changes within the profession. Metrics include the number of audits performed and the success rate.

In cooperation with the Department of Health (DOH), certain health care professionals licensed by DOS complete a survey upon renewal. The surveys are designed by DOH to collect information on the characteristics and distribution of the health care workforce in the Commonwealth. Workforce Reports inform policymakers, associations, and other stakeholders regarding the supply of health professionals. Over 355,000 Pennsylvania licensed dentists, dental hygienists, nurses, and physicians are surveyed. In addition, the Board of Nursing shares Annual Report data from its 150 Board-approved nursing programs with DOH. A summary report is developed by DOH and is utilized by policymakers, schools, associations, and other states. The Nurse and Health Licensing Divisions are required by federal law to electronically report disciplinary data to the National Practitioner Data Bank. The Data Bank is accessed by entities authorized by federal law. The Boards and Commissions within BPOA share data with associations and national councils upon request.

Outcome Monitoring

Bureau of Commissions, Elections, and Legislation. The bureau monitors the percentage of campaign finance reports and lobbying disclosure reports filed online versus on paper to determine the effectiveness of its voluntary online filing options. The bureau also monitors agency-based voter registration application activity reported by various state agencies by county and region.

Bureau of Corporations and Charitable Organizations (BCCO). The volume of filings processed is measured daily. Reports capture each section of the defined filing process. The volume of work by each employee can be monitored during their working hours. The bureau also identifies when challenges are present for a specific form type or individual. The bureau's executive office receives a daily report on processing time, and the bureau has the ability to report the total number of filings in a given timeframe.

Professional Health Monitoring Program (PHMP). Reporting of outcomes is done on an as-requested basis, not on a routine basis to any agencies, organizations, or boards. Successful completion of the program by licensees is one of the metrics for determining the effectiveness of PHMP. Since the program's primary mission is to protect the public, PHMP has identified the need to collect and analyze data on the following issues: timeliness of processing new referrals to PHMP, and actions taken by PHMP to address violations, including remanding files to the DOS Legal Office for the initiation of disciplinary action against licensees for violating a board agreement or order.

Targeted Evaluation

Bureau of Professional and Occupational Affairs (BPOA). BPOA conducts evaluations of similar programs to narrow scopes of work and training, while allowing a larger segment of the community to be eligible to practice a profession. For example, the bureau might evaluate the training and skill sets for barbers and cosmetologists to determine whether there is an opportunity to reduce the amount of red tape to obtain a license. Another type of evaluation might compare distance learning to brick-and-mortar institutions on the basis of curriculum and learning outcomes.¹⁹⁴

Juvenile Court Judges' Commission (JCJC)

JCJC has mutual cooperation with PA Commission on Crime and Delinquency (PCCD), PA Council of Chief Juvenile Probation Officers (PCCJPO), and Evidence-Based Prevention & Intervention Support Center (EPISCenter). JCJC implements multiple programs to collect data and improve its programs' effectiveness. Standardized Program Evaluation Protocol (SPEP) is responsible for data rating system that determines type of effective intervention in reducing recidivism of Juvenile offenders. Also, it creates a metric by assigning points to programs based on their similar characteristics with other programs in research studies. Moreover, JCJC provides

¹⁹⁴ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

data to several universities and organizations to analyze the performance and establish science-based plans.

JCJC grants funding to all county Juvenile Probation Services (JPS) based on several conditions. First, all county JPS are expected to implement evidence-based practices in their programs. Second, JPS are expected to improve their evidence-based performances and efficiency. Third, JCJC requires the implementation of the Youth Level of Service (YLS) risk/needs assessment instrument. Fourth, each program must develop case plans that identify risks and meet the needs of each juvenile in its care.

To further improve data collection and case management, JCJC developed the Pennsylvania Juvenile Case Management System (PaJCMS). This serves as the statewide juvenile court data collection and case management system. It is actively utilized in all 67 counties to maintain electronic records of juvenile offenders, process allegations of delinquency, and monitor compliance. Furthermore, JCJC publishes annual reports using the data collected from its various programs. Those reports include overviews of juvenile court dispositions, outcome measures to community protection, and analysis to determine the impact of juvenile justice system on reducing recidivism.

Program Assessment

Pennsylvania's Juvenile Justice System Enhancement Strategy (JJSES) Statement of Purpose is as follows: "We dedicate ourselves to working in partnership to enhance the capacity of Pennsylvania's juvenile justice system to achieve its balanced and restorative justice mission by:

- Employing evidence-based practices with fidelity at every stage of the juvenile justice process;
- Collecting and analyzing the data necessary to measure the results of these efforts; and, with this knowledge,
- Striving to continuously improve the quality of our decisions, services, and programs."

In partnership with PCCD, Pennsylvania Council of Chief Juvenile Probation Officers (PCCJPO) and the EPISCenter at Penn State University, the JCJC supports the implementation of the SPEP, an activity of the JJSES. The SPEP is a validated, data-driven rating system for determining how well a program matches what research shows is effective for that particular type of intervention in reducing the recidivism of juvenile offenders. More specifically, the SPEP creates a metric by assigning points to programs according to how closely their characteristics match those associated with similar programs shown, in research studies, to have the best recidivism outcomes.

The JCJC annually publishes various reports including a recidivism report that highlights how juvenile offender characteristics and juvenile recidivism trends in Pennsylvania have changed since 2007.

Budget Development

The JCJC has conditioned eligibility for all county Juvenile Probation Services grant funding upon implementation of evidence-based practices including:

- Implementation of the YLS risk/need assessment instrument;
- Development of recommendations to the court based on the YLS results, including the identified risks and needs of each juvenile; and
- Development and implementation of case plans that target services to meet the identified risks and needs of each juvenile.

Implementation Oversight

The JCJC is coordinating the ongoing implementation of the JJSES, which is designed to enhance the capacity of Pennsylvania's juvenile justice system to achieve its balanced and restorative justice mission by employing evidence-based practices.

The JCJC mandates that all counties participating in Juvenile Probation Services grant funding report on the implementation, sustainability, and continuous quality improvements of evidence-based practices and the activities contained within the JJSES.

The JCJC staff provides consultant services to the Commonwealth's juvenile probation departments on juvenile justice matters, including the efficiency of administrative procedures, the development of programs, and legal and legislative interpretation. Juvenile Court Consultants also conduct annual court service visits with all Administrative Juvenile Court Judges and Chief Juvenile Probation Officers. The court service visit provides JCJC staff the opportunity to review department policies and practices being utilized by juvenile probation departments. Additionally, JCJC staff members have the opportunity to observe juvenile court proceedings to witness the application of the Pennsylvania Rules of Juvenile Court Procedure and best practices related to the conduct of these proceedings.

The JCJC develops and maintains the Pennsylvania Case Management System (PaJCMS), which serves as the statewide juvenile court data collection and case management system. PaJCMS is actively utilized in all 67 counties to maintain electronic records of juvenile offenders, process allegations of delinquency, monitor compliance with conditions of supervision, and maintain juvenile-specific information. Various reports, dashboards, and other quality assurance measures are also contained within the PaJCMS, and the system provides data for virtually all juvenile justice related research.

Outcome Monitoring

The JCJC publishes an annual report that provides an overview of juvenile court dispositions in Pennsylvania. Dispositional information related to placements, offenses, demographics, and dispositional reviews are included in the report, as is information related to the use of secure detention.

The JCJC annually reports outcome measures specific to community protection, competency development, and accountability measures.

The JCJC's annual publication of the recidivism report details the outcomes of youth who had a case closed from a juvenile probation department in the Commonwealth in a particular cohort year.

The JCJC completed an analysis to determine the impact of Pennsylvania's juvenile justice system on reducing delinquent youth's risk to reoffend. The outcome data indicate that the overwhelming majority of youth (approximately 80 percent) had a lower risk score at case closure than at the time of initial assessment. A full report is detailed in the JCJC's publication, "Risk Reduction in Pennsylvania's Juvenile Justice System: An Analysis of Youth Level of Service (YLS) Assessments."

The JCJC maintains digital dashboards to each of the 67 counties to provide real-time, local and statewide performance indicators and measures.

Targeted Evaluation

JCJC's Standardized Program Evaluation Protocol (SPEP) developed from a partnership of the courts, probation system, and service providers aimed at reducing recidivism. SPEP includes a "Life Cycle" that outlines program expectations for youth involved in Pennsylvania's juvenile justice system. The statement of purpose for this performance improvement process is as follows:

- Strives to improve the quality of service delivery to juvenile justice involved youth;
- Assists in the matching of the right service to the right youth for the right amount of time;
- Focuses on services for moderate- to very high-risk youth;
- Engages service providers and juvenile probation in a collaborative effort to improve services through the flexible development of a performance improvement plan; and recognizes that locally developed programs and their services can be equally effective as those supported by research.

To ensure service providers for juveniles understand the special circumstances leading to juvenile offending, they must become versed in evidence-based practices and work collaboratively with juvenile probation departments to develop treatment methods and services; service contracts for evidence-based practices are an integral tool to this end.

JCJC's comprehensive data integrity is evidenced by ongoing and continued statewide and national requests for participation in data-related research projects. Throughout 2017, these projects included, but were not limited to:

- Providing educational statistics and YLS data to Geoffrey Biringer from the Pennsylvania Human Relations Commission to use in advocating for funding for educational services in an attempt to break the “school to prison pipeline.”
- Participating in “Results First: Cost Benefit Analysis of Juvenile Delinquency Programs.” This project seeks to determine the return on investment for certain juvenile justice programs. The role of the JCJC was to calculate a baseline five-year recidivism rate of delinquent youth and map a trajectory of youth through the system based on the offense they committed.
- Providing data to the National Center for Juvenile Justice to assist Dr. Jennifer Peck of the University of Florida and associates in examining the various roles gender, age, and race play in transferring youth to criminal court.
- Continuing work with Dr. Ed Mulvey and colleagues from the University of Pittsburgh in the Second Chance Act. Most recently, JCJC staff assisted in inputting data collected from youth surveys into an SPSS database. In 2018, JCJC will be providing additional data, including recidivism outcomes, for this project.
- Assisting Research for Action by providing recidivism and other juvenile justice outcomes in an effort to determine the outcomes of youth who were discharged from PACTT-affiliated programs.
- JCJC and five local juvenile probation departments assisted Dr. Joel Miller and Dr. Carrie Maloney as they completed qualitative and quantitative research on the operationalization of risk, need, and responsivity in local policy. Their findings were presented at The American Society of Criminology (ASC) in November 2017 and their final report is being submitted for publication.¹⁹⁵

Pennsylvania Board of Probation and Parole (PBPP)

PBPP is responsible for evaluating the cases of state prison inmates who are eligible for parole. In making parole decisions, PBPP follows specific guidelines that have been developed over many years through extensive empirical research conducted by nationally recognized researchers. Further, PBPP officers and staff undergo regular vigorous training in how best to apply the guidelines to individual inmates and parolees.

For each inmate being evaluated, the board of nine members begins by categorizing the inmate’s current offense (Group one: Murderers and Sex offenders; Group Two: other violent inmates; Group Three: Non-violent inmates; and Group Four: Recidivism Risk Reduction Incentive). For inmates assigned to group one, the majority of the board must vote in favor for an inmate to be paroled. At least two decision makers must vote in favor for those candidates assigned to the second group. The third group requires one board member and a hearing examiner to vote in his or her favor. The fourth group includes inmates who are eligible for parole if a hearing examiner votes in favor of parole.

¹⁹⁵ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

Upon release, parolees are required to adhere to PBPP guidelines and conditions. Offenders are required to report to their assigned agents, live at approved residences, maintain regular contact with assigned agents, and comply with all laws, restrictions on firearms and drugs, and payment of fines and court costs. Additional special conditions may be leveled, depending on the crime.

Part of the PBPP's role is to assess the risk each paroled offender would pose to society. Two types of risk factor, static and dynamic, are considered. Static risk factors include assessments of the offender's background, such as age, number of past offenses, disabilities, victimology, race, and gender. Dynamic risk factors include the offender's conditions, treatment, counseling, education, and programs participation.

PBPP uses the Levels of Service Inventory Revised (LSI-R) in determining risk factors. LSI-R contains 54 questions, including ten criminogenic categories that are part of the parole interview process.

Pennsylvania Commission on Crime and Delinquency (PCCD)

PCCD is responsible for enhancing coordination and planning within Pennsylvania's juvenile justice systems with the goal of increasing the safety of the community and promoting justice for all citizens. In March 2017, the PCCD partnered with the Pew-MacArthur Results First Initiative to implement the Results First approach for a subset of state and grant-funded programs. Using a pilot approach, the commission will build and use Results First tools to assess 15 currently funded programs that cross multiple policy areas, including child welfare, juvenile justice, and behavioral health. The commission plans to use the cost-benefit model to demonstrate the economic impact of investing in evidence-based prevention and intervention programs and secure future funding to expand these programs across the state. Results show marked improvement in offenders' conditions and reductions in the recidivism rate.

The programs are reevaluated on a regular basis to revise and improve their practices based on effectiveness and efficiency. All programs funded by PCCD are required to submit performance measures quarterly and annual reports that track their performance.

The following programs are among those funded by PCCD:

- Juvenile Justice Systems Enhancement Strategy (JJSES)
- Standardized Program Evaluation Protocol (SPEP)
- Criminal Justice Advisory Boards (CJABs)

Program Assessment

County Probation and Parole Evidence-based Programs. PCCD is partnered with the County Chief Adult Probation and Parole Officers Association, the County Commissioners Association, the Department of Corrections, and the Board of Probation and Parole to assist the counties in advancing evidence-based programs in county probation. Justice Reinvestment Initiative (JRI) funding has been used to support the implementation and continued use of the Ohio

Risk Assessment System (ORAS), the Level of Service/Case Management Inventory (LS/CMI), and/or the Level of Service Inventory-Revised (LSI/R) assessment to support county probation improvements and other initiatives aligned with the County Chief Adult Probation and Parole Officers Association Evidence-based Strategic Plan.

PCCD evaluates outcome measures collected from counties participating in state-funded county intermediate punishment programs (i.e., alternative sentencing for non-violent offenders with substance abuse and/or mental illness). This includes assessing the recidivism rate of offenders participating in the program. County intermediate punishment's recidivism rate shows better results for offenders who participate in the program versus similar offenders who do not.¹⁹⁶

PCCD uses state Violence and Delinquency Prevention Program (VDPP) funding to support grants to implement a set list of 13 evidence-based prevention programs. These programs are selected based on the level of evidence as to their effectiveness, including using the Washington State Institute for Public Policy (WSIPP) return on investment data. The programs eligible for funding are reviewed on a regular basis to ensure that the supporting evidence still shows quality outcomes for youth and/or families.

PCCD is leading one of the Governor's GO-TIME initiatives that is looking to improve coordination of prevention program funding across state agencies, increasing the use of effective programs while reducing the use of programs that have no effect or that can actually do harm to participants, improving the collection and quality of outcome data from funded programs, and helping communities to braid various funding streams for sustainability purposes. The "Cross Systems Prevention Workgroup" is conducting an inventory of what programs are being funded, where, and via what funding sources. This work will allow for the better use of state funds by avoiding duplication of services and measuring the impact of programs funded with state or federal dollars.

PCCD supports a grant, using federal funds, to the PA Council of Chief Juvenile Probation Officers to advance the implementation and maintenance of evidence-based practices within county juvenile probation departments and juvenile courts. This is accomplished under a framework known as the Juvenile Justice System Enhancement Strategy (JJSES), which builds upon reforms initiated in Pennsylvania during the MacArthur Foundation's Models for Change Juvenile Justice Reform Initiative. The JJSES framework includes four stages for comprehensive juvenile justice system reform with specific evidence-based tools, programs, and processes incorporated into each. The Youth Level of Service/Case Management Inventory (YLS) Risk/Need Assessment, the Pennsylvania Detention Risk Assessment Instrument (PaDRAI), and Effective Practices in Community Supervision (EPICS) Model are just three of the tools and processes being implemented through JJSES, which aims to revise and adjust activities using an evidence-based decision-making process toward the goal of improving the quality and effectiveness of the juvenile justice system.

PCCD has adopted the use of the Standardized Program Evaluation Protocol (SPEP). SPEP evaluates how well a program provided to juvenile justice involved youth (whether in a residential facility or by a community-based provider) adheres to the best practices of the model. A quality

¹⁹⁶ (Results at: http://www.pccd.pa.gov/Documents/RIPvsINC_Infograph.pdf).

improvement plan is prepared, then the SPEP team returns to the site to assess how well the program has improved its service delivery. By incorporating these best practices in programs, better results can be seen in the juveniles, compared to programs that did not go through the Quality Improvement process.

Budget Development

County Probation and Parole Evidence-based Programs. See “Program Assessment” above. JRI funding was used to support these efforts. Phase 2 JRI funding could be used to continue these projects. Additionally, PCCD allocated \$1 million from the federal Edward Byrne Memorial Justice Assistance Grant Program to provide support for the statewide implementation of evidence-based programs.¹⁹⁷

County intermediate punishment’s (CIP’s) recidivism rate shows better results for offenders who participate in the program versus similar offenders who do not.¹⁹⁸ PCCD has issued a report to the General Assembly to support additional CIP funding. In addition, the report has been used to support several Phase 2 JRI reforms, which would have a savings impact on the budget if implemented by the General Assembly.

Criminal Justice Advisory Boards (CJABs). County CJABs are groups of top-level county officials that address justice issues from a systemic and policy-level perspective. Currently there are 66 established CJABs in the Commonwealth. PCCD partners with the 66 CJABs to assist in the implementation of evidence- based programs and policies that improve the quality of justice in these counties. In 2015, PCCD received the President’s Award for Excellence in Justice Policy from the National Criminal Justice Association in recognition of policies—including CJABs—that support the implementation of community-based evidence-based programs that deliver safety and justice to the state’s residents. CJAB is a state initiative that PCCD supported as a means to improve communications, strategic planning, and funding decisions at the county level. PCCD uses CJABs extensively in developing ideas for grant proposals that will have the maximum impact on the state’s current criminal justice challenges.

PCCD’s Office of Juvenile Justice and Delinquency Prevention (OJJDP) is currently participating in a pilot of the Pew Results First model, examining the Pennsylvania-specific return on investment for the list of programs described above under “Program Assessment.” This pilot is looking at the savings to the juvenile and adult criminal justice systems by preventing youth from engaging in delinquent behaviors, improving family management, and increasing academic achievement.

The JJSES Leadership Team meets bimonthly to address gaps and future needs, as well as the budgeting to support subsequent grant awards to the Council of Chief Juvenile Probation Officers. The successes achieved through the project are evident in the significant decreases in youth arrests, delinquency placements, and detention admissions, as well as increases in youth being diverted and decreases in youth recidivism. The project’s high returns on public investment

¹⁹⁷ The Edward Byrne Memorial Justice Assistance Grant Program is administered by the U.S. Department of Justice, Office of Justice Programs. <https://www.bja.gov/jag/>, accessed July 6, 2018.

¹⁹⁸ (Results at: http://www.pccd.pa.gov/Documents/RIPvsINC_Infograph.pdf)

indicate to PCCD that the project takes the system in the right direction and merits continued funding.

OJJDP made a policy decision to use limited state VDPP funds for only those programs that have proven to be effective through rigorous research and evaluations. A set list of eligible costs is provided to applicants so that they can budget for those items that are necessary for successful implementation, while having extraneous items removed from the requested budget. This allows the prevention dollars to stretch further to provide early prevention to a greater number of Pennsylvania youth and families.

Implementation Oversight

PCCD is a partner in all county-related efforts, either through work groups formed with related associations, or in direct monitoring of how counties expend their approved grant funding. In addition, PCCD is responsible for reviewing and approving all county intermediate punishment plans in advance of funding, per statutory responsibility.

PCCD's Office of Juvenile Justice and Delinquency Prevention, in coordination with the EPISCenter, will lead the Results First pilot project. It will receive data collection support from the JCJC, the Department of Corrections, and Penn State's Edna Bennett Pierce Prevention Research Center. This cross-entity collaboration will compile and analyze data to develop the Pennsylvania Results First model.

The EPISCenter at Penn State has been funded since 2008 to improve the quality of prevention and intervention programs in Pennsylvania. For prevention, EPISCenter provided detailed technical assistance to funded sites on implementation quality, data collection and reporting, and planning for sustainability of the program after the end of their grant award. For intervention, EPISCenter is leading the expansion of the SPEP in the Commonwealth. Its staff conducts the initial SPEP assessment of providers, provides quality improvement recommendations, returns to the providers to perform a reassessment, and calculates the appropriate SPEP score for the initial and reassessments, which are posted on PCCD's website. All evidence-based programs funded using VDPP funds are required to have the fidelity of their implementation assessed by the program developer in the second year of their grant to ensure that the site is implementing the program as designed and evaluated.

The JJSES project is under the direct oversight of a JJSES Leadership Team, whose membership includes representation from the PCCD Office of Juvenile Justice and Delinquency Prevention and its advisory committee, the Juvenile Justice and Delinquency Prevention Committee. The JJSES project incorporates training and technical assistance to juvenile probation officers and juvenile court judges in every aspect of evidence-based programming and practices being introduced to the field. To ensure that all initiatives maintain fidelity to the models as designed, JJSES provides training-for-trainers sessions to juvenile probation officers showing promise in any of the evidence-based programs or practices. This is a commitment to Continuous Quality Improvement (CQI) and builds in-state capacity plus sustainability for each of the initiatives.

Also, PCCD's Office of Juvenile Justice and Delinquency Prevention manages the grant award to the Pennsylvania Council of Chief Juvenile Probation Officers. Beyond the quarterly reports for fiscal and programmatic progress that are submitted by the Council to PCCD, the agency also has email and phone communication with the individual that is the Project Director, and annual on-site monitoring visits with the Project Director, all of which ensure that the project is achieving the desired results.

Outcome Monitoring

All programs funded by PCCD are required to submit performance measures quarterly. Measures are dependent upon the program, the funding used, and any additional federal or state program or funding requirements. Outcomes are then used to develop annual reports for the agency and to track the effectiveness of programs.

All evidence-based programs funded by VDPP, and any evidence-based treatment programs supported by Endowment Act funds for child victims of sexual abuse, must submit a performance measure spreadsheet as part of their quarterly reports. These spreadsheets require grantees to enter the pre/post data from their program, and the spreadsheet then automatically tracks the outcomes to ensure comparability across sites implementing the same program.

Targeted Evaluation

Due to funding limitations, it has been difficult for PCCD to implement its own targeted evaluations of programs through rigorous academic research. While federal funding, when available, has been used in the past to conduct evaluations of specific programs and processes (e.g., Batterer's Intervention Programming, Pre-Trial Assessments), a typical study conducted in partnership with a university costs \$250,000 or more and may span several years. As such, PCCD has mainly chosen to focus its efforts on promoting programs and practices that meet the standard of being evidence-based.

Crisis Intervention Training (CIT). The CIT research project is being conducted in collaboration with the University of Pittsburgh. The two-year project is an evaluation of CIT policing programs in multiple Pennsylvania counties. Preliminary findings show that CIT makes a difference in officer behavior and outcomes for mental health clients. PCCD intends to use this information to guide county and state policies and procedures.

PCCD has partnered with the University of Pittsburgh to conduct a validation study of SPEP. The major goals of this study are to examine the relationship between performance according to the SPEP scores and recidivism outcomes, and to identify ways to focus the JJSES efforts. PCCD will examine if there is a cut-off point (threshold score) at which higher scores translate into lower recidivism rates on a consistent basis.

The Prevention Research Center (PRC) at Penn State is conducting an evaluation of the effectiveness of the Communities That Care (CTC) model that PCCD has adopted in Pennsylvania. This is a replication study investigating the public health effect of CTC over 11 years in Pennsylvania. CTC is an operating system that brings together stakeholders in a community to analyze data to determine needs, select a program that has been shown to be effective in addressing

those needs, and determine the impact of that program on the community. This study is comparing communities that are using CTC to communities with near identical demographic and other statistical traits based on the results of the Pennsylvania Youth Survey (PAYS), to determine if the youth in the CTC communities show lower levels of problem behaviors and higher levels of academic and personal success. The preliminary results show lower rates of drug use and delinquency in CTC communities.

PCCD's use of EBDM allows it to allocate limited funding to those programs that are the most promising. For example, VDPP funds support only those programs with a large amount of evidence as to their effectiveness. Rather than allowing new or untested programs to be eligible for funding, PCCD looks to replicate proven models to gain their benefits for the Commonwealth. According to PCCD's 2015-2016 Annual Report, the Commonwealth saved over \$4 dollars for each dollar spent on funding CTC. In addition, youth and family outcomes attributed to the VDPP state appropriation include a 51 percent increase in school performance, 27 percent reduction in antisocial behavior, and a 76 percent increase in family cooperation.¹⁹⁹

Pennsylvania Council on the Arts (PCA)

PCA is composed of 15 private citizens and four legislators who work to provide grants and resources to Pennsylvania arts and cultural activities and organizations. PCA collects data as part of its process for awarding grants to the Pennsylvania's nonprofit arts and culture sector. The data collection includes data about financial information, jobs, and cultural facilities. The council evaluates the applications based on its "Access to the Arts" criteria. These criteria measure the effectiveness of public arts programs using the following measures:

- Demonstrated effectiveness/results in serving the general community,
- Degree of success in reaching new and non-traditional audiences,
- Effectiveness of marketing and promoting activities,
- Evidence of effectiveness of educational activities.

Performance reviews are conducted by an advisory panel composed of members who are not state employees. The panel scores each applicant's performance and presents a report to PCA for approval.

PCA's own long-range strategic planning process is on a five year cycle and includes data review as a significant component. PCA uses a number of planning processes and information sources that support EBDM.

- Grant applications are reviewed based on applicants' past performance;
- Advisory panel comments are shared with applicants and grantees, upon request;

¹⁹⁹ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

- Interim applications and final reports are submitted by grantees providing PCA updates on any major changes to the applicant and an assurance that its funded activities are carried out in accordance to the PCA’s grant award agreement; and
- Arts in Education partners work with teachers to develop an evaluation instrument to measure student outcomes from teaching artist residencies; PCA’s Arts in Education and Preserving Diverse Cultures programs have contracted with outside evaluators for several years to ensure that they provide optimal consulting and technical support to grantees.

Data from the PCA is shared with the National Assembly of State Arts Agencies and the national Arts and Economic Prosperity Study, which allow comparisons and targeted evaluations with other states on measures of state and regional outcomes, industry expenditures, jobs, and revenue generated.

Other PCA initiatives that exhibit the use of EBDM include the Entry Track/Project Stream programs, which review new arts programs and activities on an annual basis. Those programs that meet PCA’s criteria are afforded access to on-going, multi-year application processes for funding.

Program Assessment

The PCA collects substantial data through its grant making to the nonprofit arts and culture sector, including but not limited to financial information, jobs, cultural facilities, audiences, performance, and governance. Almost 70 percent of PCA grant applications are reviewed on past performance, which is quantifiable, measurable, and verifiable by audited financials, when applicable. Panels review applications based on published criteria that include financial size, quality of management, quality of arts service/program, and how they ensure the public’s access to the arts.

Panel recommendations in the form of assessment scores by criteria are then presented to PCA for final review and approval. It is important to emphasize the PCA criteria of “Access to the Arts” in the context of “effectiveness of public programs.” “Access to the Arts” evaluates applications using the following measures:

- Demonstrated effectiveness/results in serving the audience, customers, clients, and the general community;
- Degree of success in reaching new and non-traditional audiences, customers, and clients;
- Effectiveness/results in marketing and promoting activities;
- Demonstrated results in reaching underserved communities and evidence of services or accommodations for persons with disabilities; and
- Evidence of effectiveness/results of educational activities.

Applications are supported by data from the PCA Funder Report generated by DataArts, an outside nonprofit organization developed by a public-private partnership between the PCA and private cultural funders that now serves multiple state arts agencies nationally. The PCA Funder Report gives a detailed explanation of the finances and activities of applicants. It helps the PCA bring transparency, accountability, and equity to its grant making, provides access to other funders, and is available to researchers.

Arts in Education grants have oversight and evaluation from participants, administrators, and teachers built into the funded activities.

Budget Development

PCA's Council develops its budget and funding calculations every year based on review of its data and priorities established in the goals and objectives of its long-range strategic plan.

To ensure a high return on investment of public funds for the citizens of Pennsylvania, the multi-layered approach to EBDM enables the agency to provide ongoing and reliable financial support from the Commonwealth to those with the best demonstrated capacity to deliver quality public programs.

Additional categories of PCA support are approved by the Council for statewide programs that demonstrate the ability to leverage non-state funds, respond to demonstrated community needs, and/or seek solutions to challenges:

- *Arts in Education and Pennsylvania Partners in the Arts.* The Council established a goal for these initiatives to ensure statewide coverage with at least one grant or service provided to every county;
- *Preserving Diverse Cultures.* This program provides grants to build capacity of organizations in culturally-specific communities.

Implementation Oversight

PCA's oversight of funded programs includes Council review of final reports from grantees along with PCA staff reviews and reports. Other significant components of oversight protocols include site visits, semi-annual meetings with stakeholders such as Pennsylvania Partners in the Arts and Arts in Education. Importantly, PCA seeks evaluations done by outside consultants to provide information and evidence for its EBDM.

Outcome Monitoring

PCA's long-range strategic planning process (five-year cycle) incorporates data review. Review of applications include evaluation of applicants' past performance. Both interim applications and final reports are submitted by grantees to provide PCA with updates on major changes and an assurance that funded activities are carried out in accordance to the PCA's grant award agreement.

PCA provides reports to the National Endowment for the Arts (NEA) every three years. These reports, which support PCAs funding requests, are competitively reviewed by panels of federal and state arts agencies' staff to assess performance and determine funding. PCA's Funder Report also supports grant applicants by utilizing data generated by DataArts, a nationally recognized nonprofit organization developed by a public-private partnership between the PCA and private cultural funders. Importantly, while the Funder Report assesses the finances and activities of grant applicants, it creates opportunities for other funders and researchers to access the data. The data collected and analyzed by PCA are shared with other states to compare results and to seek further improvements in an iterative process.

Every three years statewide NEA applications for federal funding are submitted and are competitively reviewed by panels of federal and state arts agency staff to assess performance and determine funding levels. Data from the PCA is also shared with the National Assembly of State Arts Agencies to allow comparison with other states. PCA data are also provided to the national Arts and Economic Prosperity project to study of the economic impact of nonprofit arts and cultural organizations and their audiences. The study includes state and regional measures and outcomes, including industry expenditures, jobs, and revenue generated.

Targeted Evaluation

Entry Track/Project Stream. The primary purpose of these two programs is to review new (or new to the PCA) arts programs and activities on an annual basis before allowing access to on-going, multi-year application processes for funding. Arts in Education has contracted with an outside educational evaluator for several years. Preserving Diverse Cultures has contracted with outside evaluators and provides consulting and technical support to division grantees.²⁰⁰

Pennsylvania Emergency Management Agency (PEMA)

PEMA's mission is to assist the public in any disaster. The agency is responsible for coordinating state and federal government resources in all phases, from preparing for to responding to disasters in Pennsylvania. Through accreditation processes, regular assessments, reviews, and evaluations, PEMA applies EBDM to its regular program and policy functions. For example, all programs are evaluated and accredited for effectiveness on five-year bases. PEMA utilizes the Threat and Hazard Identification and Risk Assessment (THIRA) and State Preparedness Report (SPR) to direct resources to areas of need.

Program Assessment

Emergency Management Accreditation Program (EMAP). All emergency management programs and processes at the agency are submitted for five-year accreditation based on their effectiveness.

²⁰⁰ Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

State Fire Academy. The International Fire Service Accreditation Congress (IFSAC) is one of PEMA's accrediting agencies that gives the Pennsylvania State Fire Academy (PSFA) and the Office of the State Fire Commissioner (OSFC) the credential to be able to conduct certification testing for fire fighters as written in the Fire Commissioner Act.²⁰¹

Budget Development

PEMA uses the Threat and Hazard Identification and Risk Assessment (THIRA) and the State Preparedness Report (SPR) to develop and prepare its annual budget for the General Assembly's consideration. The two tools provide evidence for the EBDM process that directs funding to areas of need. Similarly, a PEMA review committee determines which hazardous materials projects to fund and the dollar values for each project.

Implementation Oversight

- 9-1-1 legislation requires monitoring of program efficiencies and funding objectives.
- Equipment and equipment fund tracking are reviewed for eligibility.
- THIRA process and preparation of SPR.

Outcome Monitoring

- Federal requirements mandate policy, implementation, and follow-up monitoring.
- Hazard mitigation of flood-zone purchases

Targeted Evaluation

- Programs are evaluated based on impact to THIRA and SPR.
- Evaluation for alignment with EMAP standards²⁰²

Pennsylvania Liquor Control Board (PLCB)

The PLCB is unique among state agencies, as it serves not only as the state's regulator of the beverage alcohol system but also as a business operating as the wholesaler and retailer of liquor (wine and spirits) in the Commonwealth. The complexities associated with regulating more than 20,000 licensees and permit holders, operating more than 600 retail stores, and serving as the wholesaler to licensees and permit holders across the Commonwealth present the agency with unique challenges in developing initiatives and making decisions.

Decisions related to business functions of the agency are, very simply, based on profitability. Act 39 of 2016 made sweeping changes to the Liquor Code and required the PLCB

²⁰¹ Act of Nov. 13, 1995, (P.L. 604, No. 61), known as the "State Fire Commissioner Act."

²⁰² Letter from Colby Clabaugh, Executive Budget Specialist, Office of Budget to Joint State Government Commission, March 15, 2018.

to make significant changes to its operational and business areas. Act 39 also made clear that the PLCB is expected to maximize Commonwealth revenue; the agency has been making historically high transfers of proceeds to the General Fund in the years since enactment. Consequently, the agency's decisions related to its business functions as a retailer and a wholesaler are predicated on financial results and, ultimately, on profitability.

Program Assessment

As the PLCB is tasked with maximizing its financial returns on behalf of all Pennsylvanians, the financial results of the retail and wholesale sale of liquor are the primary key performance indicators of the agency's business functions. The PLCB monitors its overall financial performance through monthly financial statements, prepared by the Office of the Comptroller under Generally Accepted Accounting Principles (GAAP), detailing the agency's net position, revenues, and cash flows. Financial information on operating expenses, cost of goods sold, net income, and gross sales are evaluated by management against projected goals for the fiscal year.

Financial metrics are used to assess business decisions and adjust programs depending on the fiscal impact of the expenditure. For example, Act 39 removed the restriction on the number of Fine Wine & Good Spirits stores that could be open on Sundays. Based on sales data, customer trends and other factors, the legislation represented a new opportunity to provide additional customer convenience and increase sales. As a result, the PLCB moved aggressively to open additional stores on Sundays. Since the implementation of this new program, retail operations and finance staff have tracked sales outcomes and expenses, and made corresponding adjustments in store hours where appropriate.

With regard to the PLCB's regulatory functions, licensing and alcohol education processes are monitored to ensure efficiency and effectiveness. Key performance measures include the tracking of licensing investigations, application processes, and licensing compliance complaints. Alcohol education staff tracks Responsible Alcohol Management Program (RAMP) training metrics.

The PLCB recently introduced a new statewide Alcohol Education campaign entitled, "Know When. Know How." This research-based education and prevention campaign was introduced to give parents of children eight to 12 years of age the tools, resources, and confidence they need to engage in meaningful conversations with their children about alcohol. The outcomes of the campaign will be tracked by surveying the change in parental attitudes about speaking to their children at an early age about the effects of underage drinking.

Budget Development & Implementation Oversight

Annually, the PLCB develops a fiscal year budget request for the State Stores Fund in consultation with the Governor's Budget Office, and testifies before the Pennsylvania House and Senate Appropriations Committees. The PLCB, unlike most other state agencies, is not funded by an appropriation from the General Fund. Rather, the agency is funded through an executive authorization to utilize monies from the State Stores Fund, which is primarily funded through the

sale of wine and spirits in the Commonwealth. The PLCB's budget submission outlines its strategic and programmatic priorities for the coming fiscal year and identifies key performance metrics for certain initiatives.

Agency projects are fully vetted prior to engaging in contracts and are actively managed by PLCB staff. Expenditures exceeding \$50,000 must be approved by the three-member Board. Expenditures under that threshold must be approved by the Executive Director, and, in certain circumstances, the three-member Board as well. Additionally, PLCB executive staff meets on a quarterly basis with finance staff to actively monitor office budgets and assess program performance. Weekly enterprise project status meetings track progress on major initiatives across the agency to ensure communication between PLCB business areas and to identify and resolve implementation issues.

Outcome Monitoring & Targeted Evaluation

The PLCB uses an Oracle-based Enterprise Resource Planning (ERP) operating system to produce business dashboards and track key performance indicators to actively manage expense growth, address variances in planned costs and identify potential challenges. GAAP sales data and incremental margin gains are monitored on a weekly basis to assess trends. All areas of the PLCB have numerous key metrics that are tracked on a regular basis to assess success or need for adjustment. For example, aggregate incremental margin gains are tracked on products included in flexible pricing under Act 39 and comparative metrics, such as inventory shrink and advertising effectiveness, have been developed to monitor program performance. Further, as leases expire, store locations are evaluated based on sales per square foot, net income, and demographic trends to make determinations regarding store rebranding, relocation, or closure.

With over \$2.5 billion in sales (including taxes) during Fiscal Year 2016-17, the PLCB has access to a plethora of business data. The PLCB is currently in the process of improving its data analytic capabilities through the implementation of a Microsoft Power BI platform to provide increased accessibility to data by PLCB management. This platform will allow staff to have immediate access to data that are currently housed in separate systems. Additionally, the analytic capabilities of Power BI will allow staff to easily create unique data queries that are specific to their business area. These queries would have previously required a request to the IT office to cull the pertinent information.

Within the construct of the agency's responsibilities under the Liquor Code, the PLCB works to maximize the use of agency resources through the implementation of best practices across the enterprise. As previously noted, the financial results of the agency are the critical metrics to assess the outcome of the business enterprise. In the interest of public transparency much of this information is accessible in the PLCB's annual report and through various financial documents on the PLCB's website.

Pennsylvania Turnpike Commission (PTC)

PTC is responsible for providing turnpike customers with excellent service and safety. PTC implements evidence-based decision policies by comparing program trends and past results. Through management of its Strategic Plan, PTC has incorporated a roadmap to implement EBDM, utilizing the best available research and data in critical programs areas, which determine the focus and direction of investment for PTC employees, customers, and business partners.²⁰³ Also, the commission has incorporated a strategic plan that utilizes research data in critical programs to determine the commission's focus of future investments and business. As a result, the commission is able to determine the efficiency of each program and improve inefficient programs.

PTC seeks to implement EBDM in its Strategic Driver Program by establishing the goal of a reduction in the number and severity of crashes. One of the important steps is PTC's effort to develop methodology to compare the average crash rates for the PTC roadway system to the statewide average crash rates. An important finding from the crash rate analysis indicated that a reduction in driver error provides the largest opportunity for a reduction in crashes, injuries and deaths. Consequently, the Traffic Incident & Operations Center (TEO) budget was revised to include tools to reduce driver distraction by promoting better driver awareness.

The TEO Center, along with the PTC General Consulting Engineer (GCE) will monitor the implementation of new tools designed to measure the effectiveness of safety programs and report outcome data to determine if the tools are achieving a reduction in crashes. TEO, with the support from the Engineering and Maintenance Departments, the Incident Management Committee, and the Work Zone Safety Subcommittee will use Crash Cluster Evaluation to determine which safe driving initiatives warrant continued funding.

²⁰³ Letter from Stacia A. Ritter, Director, Policy, and External Affairs, PA Turnpike Commission to Glenn Pasewicz, February 23, 2018.

States' Engagement in Evidence-based Decision Making

After the 2008 recession, EBDM became prioritized by multiple states to cut inefficient and unproductive programs and redirect resources to programs that produced successful outcomes. While states' implementation of EBDM varies, they all share the common goals of achieving cost effective and efficient policies and programs that result in positive outcomes.

The Pew-MacArthur Foundation's report, *How States Engage in Evidence-Based Policymaking*, identified the top five states engaged in EBDM as Washington, Utah, Minnesota, Connecticut, and Oregon.²⁰⁴ These states are recognized for their EBDM processes and tools used to shape public policies.

Washington

Washington, widely recognized for its comprehensive use of EBDM, relies on the Washington State Institute for Public Policy (WSIPP) to develop plans to measure the effectiveness of a number of programs, such as those connected to juvenile justice.²⁰⁵ Legislative mandates require state agencies to use EBDM tools that assist leaders in incorporating information about the effectiveness of programs in making policy decisions. WSIPP is recognized nationally and internationally for its development of a set of common definitions used to identify cost effective programs and policies. Furthermore, WSIPP developed a cutting-edge cost benefit model that calculates the return on investment and is used to help determine funding for evidence-based programs.

Utah

Utah established a system of EBDM that includes both centralized and agency-specific tools. The state Office of Management and Budget requires each agency to answer a set of questions to show how it uses evidence-based practices and research data when determining funding. The programs that fail to meet the standards must conform to receive future funding. Further, Utah's successful implementation of EBDM was supported by establishing a mandatory statewide registry of evidence-based prevention programs and evidence-based work groups that refine and plan programs based on EBDM guidelines.²⁰⁶

²⁰⁴ Elizabeth Davies, et al, "'How States Engage in Evidence-Based Policymaking,'" The Pew-MacArthur Foundation, January 2017,

http://www.pewtrusts.org/~media/assets/2017/01/how_states_engage_in_evidence_based_policymaking.pdf.

²⁰⁵ Ibid. 8.

²⁰⁶ Ibid. 9.

Minnesota

In 2011, Minnesota launched Better Government for a Better Minnesota to improve cost-effectiveness of state-funded services. Minnesota developed and expanded its policies to fully improve outcomes. For example, the state Department of Human Services maintains an evidence-based practices database, which contains data on children's mental health programs and provides decision makers with strategies for treatment and percentage of success based on individual and demographic characteristics. Further, Minnesota operates the Minnesota Correctional Program Checklist Collaborative, which utilizes the University of Cincinnati's Correctional Program Checklist tool. This tool provides an assessment of successful programs and their usage of evidence-based practices. Moreover, the state partnered with Results First to categorize programs based on evidence effectiveness and cost-benefit analysis to estimate program investment returns.²⁰⁷

Connecticut

In 2005, Connecticut engaged in Results-Based Accountability, a budgeting technique that assists policymakers in using data on program outcomes to determine their funding decisions. Through structured and organized communication tools, policymakers learn which programs are effective and successful; this information is used to make decisions regarding investing into programs. In 2015, the legislature passed a law requiring several elements of an effective evidence-based system, requiring agencies to create tiers of evidence for programs and cost-benefit analysis of all of their programs.²⁰⁸

Oregon

Oregon law requires human services agencies to spend at least 75 percent of their funds on evidence-based programs. Also, the law guides agencies to identify cost-effective programs through use of a cost-benefit tool. The state mandates progress reports to ensure agencies are staying on track. As a result, various agencies incorporated evidence-based programs. For example, Oregon Youth Authority reported targeting 89 percent of all funds to evidence-based programs.²⁰⁹

²⁰⁷ Ibid.

²⁰⁸ Ibid.

²⁰⁹ Ibid.

RECOMMENDATIONS FOR PENNSYLVANIA’S USE OF EVIDENCE-BASED DECISION MAKING

Governments that effectively respond to the needs of constituents have always made decisions based on available information. As all levels of government acquire increasingly sophisticated yet inexpensive information technology, empirical methods that normally had been in the hands of academic and scientific researchers are moving from the benchtop and into government offices. Policy makers can now and for the foreseeable future take full advantage of the opportunities presented by powerful data analytics. Consequently, the use of empirical tools similar to those employed in the classic scientific method have been bundled into a public policy decision making paradigm referred to as EBDM.

Pennsylvania is among a number of states that have embraced EBDM as the standard for policy making. Pennsylvania’s state government covers a broad array of responsibilities and obligations; each department serves a specific purpose. Although close collaboration between agencies is necessary because departments share some areas of responsibility, each agency must meet its own statutory responsibilities. EBDM provides the evidence of what works and what does not—a benefit in an era of scrutiny over government accountability. How well each agency accomplishes its mission and achieves its goals can, and should, be determined by EBDM. Importantly, the use of EBDM will help spur development of new programs and the reallocation of resources in existing programs.

For example, EBDM may lead to a better service delivery model for DHS’s Community Health Choices. It may lead to a more effective way of operating L&I’s Unemployment Compensation Service Centers. It may lead to a new approach for PDA to eradicate the Spotted Lantern Fly. Whether the focus is one program run by one bureau, or all programs administered by a full department, each must have the staff, the tools, and the resources to conduct the research necessary to engage in EBDM by decision makers.

The establishment of EBDM as the dominant decision making paradigm is not an inexpensive undertaking. State government staff will need training and knowledge of how, for example, to apply third-party research, how to gather and analyze program data, and how to interpret the results. Critically, perhaps most important, staff needs the ability to clearly, concisely, and correctly communicate the results.

Each agency that intends to conduct EBDM should have an office whose sole responsibility is EBDM. Office staff would assist the executive staff in developing Theory of Change models for each program being studied, especially new initiatives. A central office of EBDM could be established as part of the governor’s policy office and function similarly to the Office of General Counsel.

The General Assembly contains several legislative service agencies (LSAs) that have the capabilities and resources to conduct EBDM studies on its behalf. These agencies are designed to apply empirical methods to legislators' public policy proposals without political constraints, and provide ready access to expertise on whatever topics are presented. The LSAs deliver substantiated evidence to inform the legislature's deliberations to address issues and opportunities concerning the people of Pennsylvania.

Finally, state government agencies should continue to partner with organizations like the Pew-MacArthur "Results First Initiative," for guidance and resources. A prime example of collaborations between research and decision making in Pennsylvania is shown by the Pennsylvania Board of Probation and Parole, which demonstrates proven successes through its use of comprehensive and in-depth use of EBDM. Importantly, however, the agencies must not lose sight of meeting their obligations within the context of Pennsylvania's demographics, geography, politics, and socio-economic spectrum. Research cannot amend itself to the political calendar. Where necessary, studies must continue across changes in administration and be insulated from political influences.

THE GENERAL ASSEMBLY OF PENNSYLVANIA

SENATE RESOLUTION

No. 294 Session of
2015

INTRODUCED BY STEFANO, FOLMER, RAFFERTY, GREENLEAF, BROWNE,
MENSCH, VOGEL, WARD, WAGNER AND VULAKOVICH, MARCH 22, 2016

SENATOR MCGARRIGLE, INTERGOVERNMENTAL OPERATIONS, AS AMENDED,
JUNE 29, 2016

A RESOLUTION

1 Directing the Joint State Government Commission to conduct a
2 study examining the use of evidence-based decision making by
3 the Commonwealth.

4 WHEREAS, Evidence-based decision making utilizes the best
5 available research and data on programs to guide public
6 officials in determining whether to expand, eliminate or
7 modify certain government programs; and

8 WHEREAS, Evidence-based decision making has been used by the
9 Federal Government, state governments and private sector
10 companies; and

11 WHEREAS, Evidence-based decision making will allow State
12 agencies to have a better understanding of whether certain State
13 programs are an effective use of taxpayer dollars; therefore be
14 it

15 RESOLVED, That the Senate direct the Joint State Government
16 Commission to conduct a study examining the use of evidence-
17 based decision making by the Commonwealth; and be it further

18 RESOLVED, That the study include an examination of all of the

1 following:

2 (1) The use of evidence-based decision making by other
3 state governments.

4 (2) A comprehensive review and analysis of the continuum
5 of evidentiary standards used for evidence-based evaluations.

6 (3) A comprehensive review and analysis of the existing
7 evidence-based program registries currently available, the
8 standards of evidence used by each and an assessment of the
9 strengths and limitations of each, including, but not limited
10 to, the following:

11 (i) Blueprints for Healthy Youth Development.

12 (ii) The Office of Juvenile Justice and Delinquency
13 Prevention Model Programs Guide.

14 (iii) The United States Department of Justice's
15 website www.crimesolutions.gov.

16 (iv) The Substance Abuse and Mental Health Services
17 Administration National Registry of Evidence-Based
18 Programs and Practices.

19 (v) The California Evidence-Based Clearinghouse for
20 Child Welfare.

21 (vi) The Coalition for Evidence-Based Policy.

22 (vii) The Promising Practices Network.

23 (viii) The What Works Clearinghouse.

24 (4) The potential methods that the Commonwealth can use
25 to implement evidence-based decision making, including a
26 determination AND EXPLANATION of which State agencies and <--
27 programs are best suited to implement evidence-based decision
28 making.

29 (5) The amount and types of data collected by State
30 agencies and the extent to which the data is shared among

1 State agencies.

2 (6) The current methods of data protection employed by
3 State agencies and the most effective methods to protect the
4 privacy rights of the residents of this Commonwealth who
5 provide personal data to State agencies;

6 and be it further

7 RESOLVED, That the Joint State Government Commission issue a
8 report on the study to the Senate within 18 months of the
9 adoption of this resolution.