



The Nelson A. Rockefeller Center at Dartmouth College

The Center for Public Policy and the Social Sciences

The Class of 1964 Policy Research Shop

—Celebrating 10 Years of Service to New Hampshire and Vermont—

ANALYZING PROGRAM EVALUATION METHODOLOGIES

What is 'Evidence-Based' Policymaking?

Presented to the New Hampshire House Finance Committee

PRS Policy Brief 1516-04

April 12, 2016

Prepared By:

Sam Libby

Joby Bernstein

Michael Baicker

This report was written by undergraduate students at Dartmouth College under the direction of professors in the Rockefeller Center. Policy Research Shop (PRS) students produce non-partisan policy analyses and present their findings in a non-advocacy manner. The PRS is fully endowed by the Dartmouth Class of 1964 through a class gift in celebration of its 50th Anniversary given to the Center. This endowment ensures that the Policy Research Shop will continue to produce high-quality, non-partisan policy research for policymakers in New Hampshire and Vermont. The PRS was previously funded by major grants from the U.S. Department of Education, Fund for the Improvement of Post-Secondary Education (FIPSE) and from the Ford Foundation and by initial seed grants from the Surdna Foundation and the Lintilhac Foundation. Since its inception in 2005, PRS students have invested more than 50,000 hours to produce more than 130 policy briefs for policymakers in New Hampshire and Vermont.



Contact:

Nelson A. Rockefeller Center, 6082 Rockefeller Hall, Dartmouth College, Hanover, NH 03755
<http://rockefeller.dartmouth.edu/shop/> • Email: Ronald.G.Shaiko@Dartmouth.edu



TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	1
2. INTRODUCTION	1
3. DEFINING EVIDENCE-BASED	2
3.1 TERMINOLOGY	2
3.1.1 <i>What is Evidence-Based?</i>	2
3.1.2 <i>What is Cost-Effective?</i>	3
3.2 QUALITY ASSESSMENT TOOL FOR QUANTITATIVE STUDIES	3
3.3 POLICE STANDARDS AND TRAINING	8
4. PROGRAM METHODOLOGIES.....	8
4.1 PEW MACARTHUR RESULTS FIRST	9
4.2 RESULTS-BASED MANAGEMENT	10
4.3 RESULTS FOR AMERICA	11
5. LEGISLATIVE STRATEGIES	12
5.1 SUNSET REVIEW	12
5.2 STANDING COMMISSION	13
6. CONCLUSION.....	14
APPENDIX. WASHINGTON STATE	15
REFERENCES	18



1. EXECUTIVE SUMMARY

This report addresses the goal of effective and efficient governance. The New Hampshire legislature receives a number of policy proposals each session claiming to be “evidence-based” but has no program evaluation mechanism. Proposals use different definitions and frameworks, and the legislature does not currently have one standard program methodology. In this report, we outline the importance of evidence-based policymaking, define what it means to be “evidence-based” and “cost-effective,” offer a quality assessment tool for evaluating whether a proposal effectively uses evidence, and discuss possible models New Hampshire could adopt for program assessment. We present several procedures that the legislature can use to determine the results, costs, and benefits from policy proposals. We then analyze these frameworks and evaluate them in terms of their practicality, their efficiency, and their relative strength compared to programs in other states. We then present several strategies used by other states that the legislature can introduce.

2. INTRODUCTION

Every year, governments make budget and policy choices with ongoing fiscal pressure to maximize the return for constituents. Policymakers may achieve better results by using rigorous evidence that enables governments to choose, fund, and operate programs more strategically. Both the Heritage Foundation¹ and the Brookings Institution² recognize that although public policy can be informed by values, rigorous evidence may also play a significant role. Creating implementation standards for legislation can promote consistent delivery of high-quality and effective services by creating requirements for oversight and monitoring.³

Peter Schuck, a Yale Law professor, estimates that “less than 1 percent of government spending is backed by even the most basic evidence of cost-effectiveness.”⁴ Evidence-based policymaking requires policy decisions to follow rigorous impact evaluations of programs that incorporate research and academic techniques in order to parse out the true effect of the policy in question. Evidence-based policymaking is one issue policy topic that draws bipartisan support, with the goal of reducing wasteful spending, expanding innovating programs, and strengthening accountability. Identifying a social program solely through values and attempting to fix it through the creation of new programs and millions in appropriations will not necessarily alleviate the problem⁵. For instance, when 13 youth job-training programs were evaluated for cost-effectiveness, only one program was demonstrated to have a positive impact on earnings.⁶ The means to evaluate accurately the cost-effectiveness of government programs is often missing from policymaker’s “toolbox.”

Some states and the federal government are elevating evidence-based policymaking. The Office of Management and Budget (OMB) has emphasized the importance of rigorous evidence to evaluate the government’s investment and use of taxpayer funds.⁷ The



Obama administration established the White House Social and Behavioral Sciences Team to improve government efficacy.⁸ State and local governments have also begun to develop programs that link private funding to public initiatives that have produced results. New York City, Salt Lake City, New York State, and Massachusetts have developed pay-for-success models with social impact bonds, whereby private capital is raised from investors and philanthropic organizations in order to fund programs that have the potential to achieve better outcomes.⁹ For example, New York raised \$13.5 million to equip incarcerated adolescents with emotional and social skills to improve life choices after they leave jail.¹⁰ The program, funded through a social impact bond, is an example of an evidence-based employment service that has yielded financial savings to the government by reducing recidivism. Rigorous government evaluation of evidence-based policy can stimulate financing from private investors in order to produce outcomes that improve social welfare.

States across the country have initiated data inventories of the programs they operate in order to assess the available evidence and return on investment for government programs. This report will develop definitions for different categories of evidence-based programs, present a quality assessment tool for quantitative studies, outline different examples already enacted in other states and suggested by various think tanks, and present several legislative strategies that New Hampshire could copy.

3. DEFINING EVIDENCE-BASED

In this section, we define what it means to be evidence-based. We present a quantitative assessment tool that can be used to determine whether a program is evidence-based, and we walk through an example, highlighting what would need to happen in order to be considered evidence-based.

3.1 Terminology

3.1.1 What is Evidence-Based?

The term ‘evidence-based’ requires programs to prove effectiveness by generating consistent, statistically significant outcomes that improve upon initial conditions.¹¹ In order to be considered evidence-based, a program must use multiple-site randomized control trials across diverse populations.¹² Randomized control trials assign participants randomly and in equal numbers to a control or a treatment group, testing the treatment group by applying the policy, and then observing whether the outcome differs between the treatment and control group.¹³ In the absence of a randomized control trial, other research designs, including multiple variable fixed effect regressions, difference-in-differences, and regression continuity, can also be used to find unbiased estimates of program evaluation through a natural experiment.



These techniques are not a tremendous step forward for policymakers. For the past few decades, legislators have considered evidence-based standards. One part of the *No Child Left Behind Act of 2001* sets a clear standard for scientific research. In order to be “scientifically-based,” research must:

- Apply rigorous procedures
- Employ systemic, empirical methods
- Involve rigorous data analyses
- Have common measurement and observational definitions
- Accepted by a peer-reviewed journal¹⁴

Evidence-based policymaking uses similar techniques, albeit in a more targeted setting. These techniques constitute the most effective means of uncovering the true effect of policies on the target population.

3.1.2 What is Cost-Effective?

Cost-benefit analysis (CBA) is not a form of evidence-based research but rather a structured form of analysis that may guide policymakers. CBA aggregates the current and future costs of a program and discounts future costs/benefits to the net present value. If the future benefits outweigh the future costs, the program is considered cost effective. However, because a program affects multiple systems, it can be difficult to fully calculate for all possible costs and benefits into perpetuity. Due to the ability to alter assumptions and discount rates within a cost-benefit analysis, cost-effective research using CBA must be interpreted with caution.

3.2 Quality assessment tool for quantitative studies

The purpose of this tool is to assist legislators and government officials to evaluate the merits of quantitative studies on current or future programs. This tool can be used for both future program assessment and outcome monitoring, and it can judge whether a program meets standards of evidence-based policy. The tool provides information on what should be included within quantitative program assessment studies in order to assess the validity, reliability, and applicability of the research. Program assessments that receive lower scores from this tool may not be considered “evidence-based.” The tool was developed based on models presented by the Effective Public Health Practice Project¹⁵ and Research Connections.org.¹⁶

1. **Population.** Was the population *eligible* to be selected for the study include the entire population of interest? For example, are all the teenagers in New Hampshire eligible to be selected for the study (i.e., the entire population of interest)? Or, were only children in Grafton County eligible to be selected for the study (a selective subgroup)?

[2] Eligible population includes entire population of interest or a substantial portion



[1] Eligible population includes a substantial portion of the entire population of interest

[0] Population represents a limited or selective subgroup of the population of interest

[-1] Population represents an atypical subgroup of the population of interest

[-2] No description of the population

[NA] Not applicable

2. Randomized Selection. Were study participants randomly selected for the study? Or, did study participants volunteer (nonrandom)? Or, were they located through specific organizations (nonrandom) or locations (nonrandom)?

[1] Random selection

[0] Nonrandom selection

[-1] No description of the sample selection procedure

[NA] Not applicable

3. Sample Size. How many participants were selected for the study? Does the sample include enough participants from key subgroups to accurately assess subgroup differences? Does the study compare the sample size to other similar studies?

[1] Sample size larger than similar studies

[0] Sample size the same as similar studies

[-1] Sample size smaller than similar study or sample size not given

[NA] Not applicable

4. Response and Attrition Rate. What proportion of the selected sample completed the study? In longitudinal studies, what proportion of sample members participated in follow-up studies?

[2] High response and participation rate (response rate above 80%)

[1] Moderate participation rate (response rates at least 50%)

[0] Limited response and participation rate (response rate at least 30%)

[-1] Poor response rate (less than 30%)



- [-2] No information on response rate or participation rate
- [NA] Not applicable
- 5. Main Variables.** Are independent and dependent variables of interest described fully?
- [1] Accurately described and can be matched
- [0] Vague definition or cannot be matched
- [-1] No definition of main variables or concepts
- [NA] Not applicable
- 6. Randomization Methods.** Was the study described as randomized? Was the method clearly described?
- [1] Randomized study, method of randomization clearly described
- [0] Not randomized, but uses a conventional approach such as difference-in-difference or regression discontinuity
- [-1] No randomization methods described
- [NA] Not applicable
- 7. Numeric Results.** Are the means (or measurement standard of interest) and standard deviations/standard errors for all the numeric variables presented?
- [1] Means and standard deviations/standard errors presented
- [0] Means, but no standard deviations/standard errors presented
- [-1] Neither means nor standard deviations/standard errors presented
- [NA] Not applicable
- 8. Missing Data.** Are the number of cases with missing data specified? Is the statistical procedure(s) for handling missing data described?
- [1] Number of cases with missing data are specified and the strategy for handling missing data is described
- [0] Number of cases with missing data specified, but these cases are removed from the analysis
- [-1] Missing data issues not discussed



[NA] Not applicable

9. Appropriateness of Statistical Techniques. Does the study describe the statistical techniques implemented? Does the study explain why the statistical technique was chosen?

[1] Statistical techniques, reasons for choosing technique, and caveats are fully explained

[0] Statistical technique is explained, but the reasons for choosing technique are not included.

[-1] Statistical technique, reasons for choosing technique, and caveats are not explained.

[NA] Not applicable

10. Omitted Variable Bias. Could the results of the study be due to alternative explanations that are not addressed in the study? Does the study address possible sources of bias?

[1] All important explanations are included in the analysis, possible omitted variables or bias are considered

[0] Discussion about bias are omitted from the analysis

[-1] Variables and concepts might be biased in the study

[NA] Not applicable

11. Analysis of Main Effect. Are coefficients for the main effect interpreted? Are significance levels or the results of statistical tests presented?

[1] Model coefficients and significance levels for the main effects variables are presented

[0] Either model coefficients or hypothesis tests for the main effects variables are presented

[-1] Neither estimated coefficients or significance levels for the main effects variables are presented

[NA] Not applicable

12. Peer Reviewed. Was the study peer reviewed by a reliable source?



[1] The study was peer reviewed by an independent source

[-1] The study was not peer reviewed by an independent source

[NA] Not applicable

13. Balance Tests. Are differences between treatment and control groups assessed? Are possible significant decreases considered?

[1] Difference between groups is very minimal, and differences are clearly described

[0] Differences between groups is significant, but attempts to address possible bias in assumptions are addressed

[-1] No description of differences between experimental groups

[NA] Not applicable

14. Controlling. Does the study control for possible confounding variables? Examples include race, sex, family status, age, socioeconomic status, education, health status, etc.

[1] Yes, controls are fully addressed and included within the research

[0] Some controls are included or discussed

[-1] No mention of including controls

[NA] Not applicable

15. Data Collection. Were data collection methods fully described? Does the study account for possible collection errors? Was data collection shown to be valid and reliable?

[1] Yes, data collection methods fully described

[0] Data collection methods somewhat described

[-1] No mention of data collection

[NA] Not applicable



3.3 Police Standards and Training

In 2013, Governor Hassan created the Governor’s Commission on State Government Innovation, Efficiency and Transparency to bring together experts from the private and public sectors as well as nonprofits to save taxpayer dollars and improve state services.¹⁷ The final Governor’s report includes an appendix on efficiency and innovative ideas identified by state agencies.¹⁸ Although the dozens of programs included within the appendix claim to be efficiency improvements, very few of them are evidence-based. The report views programs as streamlined, cost-effective, or efficient; yet, there is no quantitative experimental evidence presented behind these claims. The following presents an example of how these programs could be assessed through our Quantitative Research Assessment Tool.

“More recently, over the last couple years, we transitioned away from laptops which were heavy and had limited battery life to iPADS which are more portable, lighter and have exceptional battery life. We have also found that the iPADS are more intuitive for the user making them easier to deploy. This transition to technology had made material delivery easier and more cost effective. Testing on the system is more immediate with results and also allows for us to perform quicker analysis of each individual question results to identify any deficiencies in material or learning delivery. While these changes don’t affect the citizens of the state directly, it does allow us to provide better service to the officers and agencies that we serve in law enforcement.”¹⁹

The following example is not an example of evidence based policy assessment. In order to examine if this program is cost-effective or evidence based, the program would need to be examined through a randomized control trial. Our quantitative assessment tool would require ensuring that the eligible population was the entire population of interest, there was random assignment for treatment and control groups, response and attrition rates were accounted for, with the main dependent variables being described. Furthermore, numeric results would need to be analyzed with standard errors, statistical significance tests, examining missing data, and accounting for omitted variables. This data analysis and assessment is very important in order to make sure that all efficiency claims have internal and external validity.

4. PROGRAM METHODOLOGIES

In this section, we have outlined some of the best practices across the country that use an evidence-based approach. These models explore different methodologies states and organizations have taken to improve program evaluation to focus on results and efficiency.



4.1 Pew MacArthur Results First

The Pew-MacArthur Results First Initiative began in 2011 as a partnership with states and counties to create an evidence-based framework with five key components for evidence-based policymaking. These components include (1) program assessment, (2) budget development, (3) implementation oversight, (4) outcome monitoring, and (5) targeted evaluation.²⁰

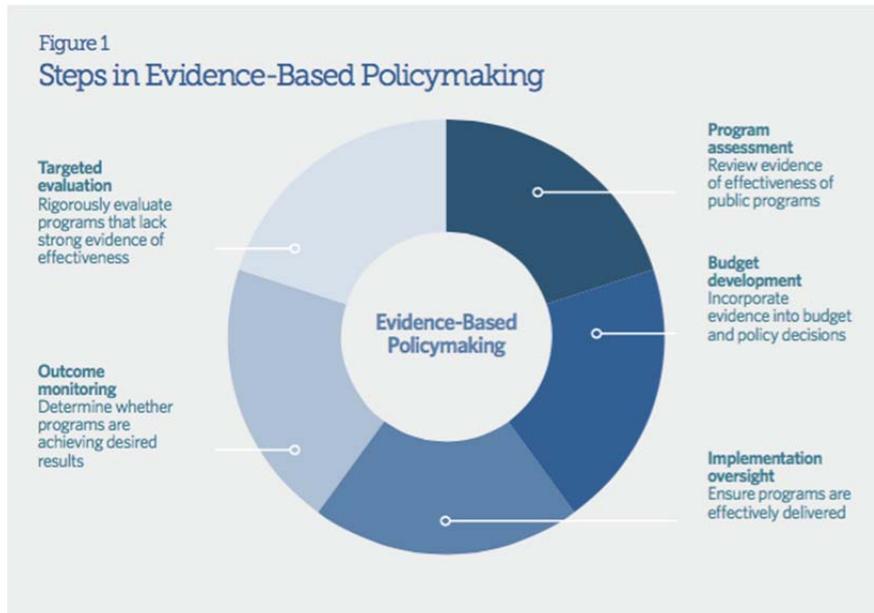


Figure 1. Defining Evidence-Based Policymaking

The Results First model is based on a methodology created by the Washington State Institute of Public Policy (WSIPP) in the early 1990s. More details about WSIPP, including full details on the model on which Results First is based, can be found in Appendix A. The model, originally designed for implementation in the criminal justice arena, focuses on creating program inventories, evaluating state data against national data, and measuring cost-effectiveness of individual programs to target those that do and do not work. It has since been standardized by Pew-MacArthur and expanded into several other policy areas across many states. Iowa, Massachusetts, New Mexico, New York, and Vermont have used this model to redirect \$81 million in program funding to sources that are more effective.

A Pew Results First study found that the number of states using evidence-based analysis for policy decisions increased 48 percent between 2008 and 2011, now totaling 22 states and 4 counties in California using the Results First model. In 2012, Washington passed legislation that created criteria to increase the number of evidence-based social programs. Mississippi enacted similar legislation in 2014 mandating that programs within state agencies be categorized based on their use of evidence.²¹ In Iowa, the Department of Corrections found that behavioral therapy programs returned \$37.70 in benefits for every



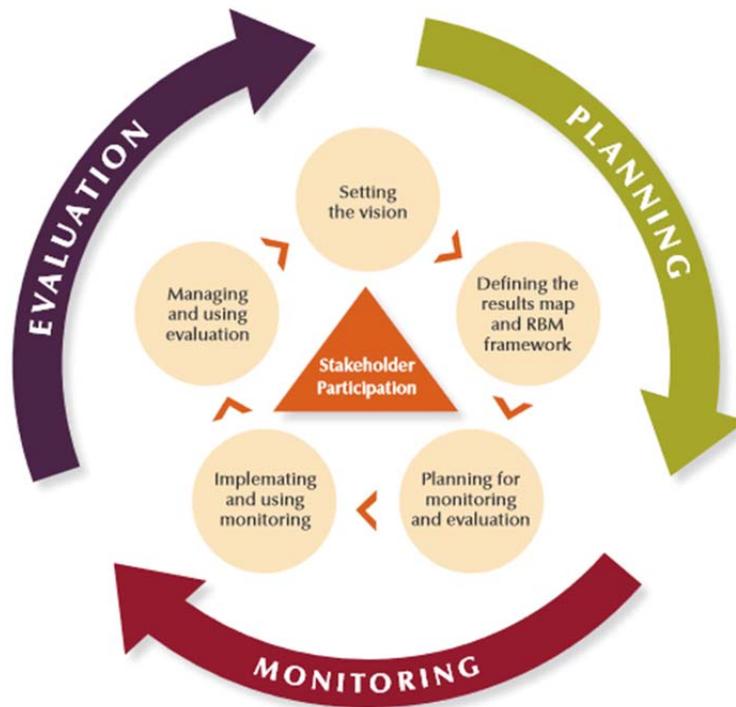
dollar spent, while correctional education programs, although effective, only returned \$2.91 in benefits for every dollar spent.²²

Pew maintains a Results First Clearinghouse Database on their website that is the primary databank used by states to compare the effectiveness of current and potential programs.²³ The database consists of program evaluations from eight separate research clearinghouses, each with its own methodologies and performance metrics for program evaluation. While this fails to provide a standardized comparison, Pew has color-coded programs into like-categories, such as “Highest rated,” “second-highest rated,” “no evidence of results,” “mixed effects,” and “negative effects,” allowing comparisons to be made across organizations.

States that join Pew’s Results First initiative receive help in creating statewide program inventories with all necessary information to compare to the existing data in their national Clearinghouse Database.²⁴ States, however, conduct analysis. While some have assigned subcommittees within the legislature for this task, most will hire one or two additional staff members dedicated to this purpose. Given the relative lack of resources within the NH legislature, this may not be a feasible option. It is for this reason that Pew grants unrestricted access to the Clearinghouse Database online. Legislators can then compare proposed programs in New Hampshire to those already listed on the database and thus evaluate whether or not they would be successful. Each clearinghouse conducts its own program evaluation, which is easily searchable on the online database.

4.2 Results-Based Management

Results-based management (RBM) is a program management technique developed by the United Nations. The RBM framework works across the entire life of the program, guiding policymakers from planning through program monitoring and evaluation. RBM depends on three principles: accountability, national ownership, and stakeholder inclusiveness. While the RBM model as dictated by the United Nations is built more for developing states, the framework itself could apply to any program implementation within the United States. Figure 2 breaks down the steps behind RBM, illustrating how to develop a program and continually evaluate its success.



Source: UNDP, *Handbook on Planning, Monitoring and Evaluating for Development Results*, 2009.

Figure 2. The RBM Life-Cycle Approach

RBM defines results by conducting a country analysis that analyzes the underlying problems and designates areas of focus. RBM establishes a results chain that establishes a clear causal link between the specific actions taken and the outcomes achieved. In doing so, RBM is able to infer causality because the approach isolates the specific actions taken as part of a program and establishes a key framework for measuring the results. RBM establishes clear output indicators, a baseline comparison, and a goal, collecting data to determine how the output resolves with the baseline and whether the goal has been met.²⁵

4.3 Results for America

Founded in 2012, Results for America is a nonprofit that aims to use best practices for policymaking. Results for America builds on progress made by the Obama Administration and Congress to fund programs that use evidence to determine which programs to fund, at what level they should be funded, and whether they should be funded in the future. The new evidence-based programs depend on incentives and evaluations to ensure a focus on results and smart policymaking. Recipients of federal grants under these programs use a tiered-evidence framework that allocates more funding to programs that expand what already works, with less money funding promising practices and the least money going to program development.²⁶ A tiered funding structure is a possible implementation method for New Hampshire moving forward, allowing the



state to pilot certain programs deemed “evidence-based” without a large initial monetary commitment.

5. LEGISLATIVE STRATEGIES

In this section, we explore a number of legislative strategies the New Hampshire legislature could introduce to implement one or more of the aforementioned methodologies.

5.1 Sunset Review

Legislators across the country use sunset provisions to increase program accountability and analyze program effectiveness. Sunset requires legislation to expire on a certain date unless the legislature takes action to renew the law.²⁷ While both states and Congress use sunset provisions, the process is more common in the states, with only nine states lacking an active sunset process. New Hampshire is one of those nine states, having repealed its sunset committee in 1986.²⁸ While bills that establish sunset committees have been introduced, such as in 2011, when members of the House proposed a joint legislative committee to review executive and judicial programs every six years, no sunset committee has been enacted.

Sunset commissions generally renew executive programs. In a study by the Mercatus Center, researchers found that 21 percent of reviews ended with a program termination; however, one state, Ohio, heavily influenced that figure, and accounting for its effect reduces the number to 11 percent in the states surveyed. Many states have seen a high return on their investment in sunset commissions, with Texas seeing a 27:1 return and Minnesota seeing a 42:1 return. In Texas, reviews over a 27-year period cost roughly \$28.6 million compared to \$783.7 million in estimated savings.²⁹

Many of the programs analyzed were state licensing boards. In July 2015, the White House released a report examining the growth of occupational licensing in the states and evaluating its economic effects. The report found that closely targeted licensing provisions can lead to greater safety and higher quality, but overly broad requirements reduce employment, artificially increase wages for licensed workers, and increase prices for consumers. Furthermore, of over 1,100 occupations regulated by at least one state, less than 60 are regulated by every state.³⁰ A sunset commission in New Hampshire could weigh the costs and benefits of policies such as licensing boards, ensuring that policies aimed to protect consumers can demonstrate results empirically.

Texas has a well-established Sunset Commission that may serve as an example for New Hampshire. Under the Texas model, a 12-member commission consisting of five members of the Senate, five members of the House, and two public members appointed by the Lieutenant Governor and the Speaker of the House, respectively, reviews every executive agency that is set to expire and recommends its continuation or abolition. The



legislature then votes on whether to extend the agency. Staff review reports, solicit public feedback, and receive input from relevant parties. The Commission then holds public hearings and publishes a staff report. Finally, the Commission holds a public hearing to vote on a recommendation.³¹ The Commission costs roughly \$2.3 million annually.³²

5.2 Standing Commission

While we have critiqued the characterization by the Governor's Commission on State Government Innovation, Efficiency and Transparency of its findings as "evidence-based," we present a standing commission as a potential policy solution that could evaluate and promote evidence-based policies. The Governor's Commission argues that despite the potential lack of political will for reform, the dividends earned will be essential to ensure long-term fiscal stability and safeguard vital executive programs in the event of revenue shortfalls. The Commission acknowledges its authority stems from the New Hampshire Constitution, as Article 8 states that "Government, therefore, should be open, accessible, accountable and responsive."³³

New Hampshire has already taken several steps to increase efficiency, including the development of Lean, a methodology that helps government employees identify and correct inefficiencies. Lean has been implemented across the New Hampshire government, from the National Guard to the executive departments.³⁴ Lean, for example, has helped the Department of Transportation save \$2 million annually in toll operations and has decreased the time between arrest and prosecution for misdemeanors by roughly 60 percent. However, while New Hampshire has established numerous other commissions to increase efficiency, there is no standing framework for reporting and communicating efficiency and innovation improvements.³⁵

The Commission's report included a number of recommendations for New Hampshire to improve its executive branch efficiency and establish dedicated funds for increased effectiveness. We have highlighted the recommendations most relevant to increasing program cost-effectiveness:

- Creating an Office of Operating Performance within the Governor's Office to oversee the executive branch. This new office would include an Office of Operating Performance that primarily monitors and drives executive branch performance, but it would lack line authority. Additionally, an Office of the Chief Operating Officer would assume operating responsibility, while the departments and agencies would retain institutional discretion. The additional level of management would streamline reporting and give the Governor additional oversight.
- Establishing clear standards for efficiency, innovation, and transparency understood by the entire executive branch. These standards would help measure performance, including the use of balanced scorecards and operating reviews.³⁶



A commission to oversee program evaluation is not a novel idea. The House of Representatives recently approved the Evidence-Based Policymaking Commission Act of 2015 (H.R. 1831), which would establish a commission in the executive branch to create a data clearinghouse for federal programs. The bill was sponsored by Rep. Paul Ryan (R-WI) and has received bipartisan support in both houses of Congress. The Act would create a federal commission consisting of three experts appointed by the President, Speaker of the House, House Minority Leader, Senate Majority Leader, and Senate Minority Leader. The Commission would have a budget of \$3 million allocated from Census Bureau funding to study federal data collecting and determine whether to establish a federal clearinghouse, with the ultimate goal of recommending legislation or administrative reform.³⁷

6. CONCLUSION

As a political branch, the New Hampshire legislature may decide how to evaluate program effectiveness and what categories to prioritize. The methodologies and strategies that we highlighted in this report are examples of best practices across the country. However, there is no panacea for policymakers to solve every policy problem before them. As a political body, the New Hampshire legislature makes tough choices about which data metrics it wants to prioritize. The methodologies and strategies we present are several tools in a larger toolkit for policymaking, and it is up to the legislature to decide which tools to prioritize. Evidence-based policymaking is one approach that New Hampshire can consider, but the legislature ultimately makes the tough decisions on how to best allocate scarce resources.

If New Hampshire chooses to pursue evidence-based policymaking, then it could establish a framework for determining which proposals fulfill the necessary criteria. The Quantitative Research Assessment Tool that we present is one way to determine whether important benchmarks have been met. We have identified other methodologies and techniques for effective policymaking that New Hampshire may also implement. As a small state with limited revenue streams, New Hampshire faces challenges identifying which programs work. By shifting to a more evidence-based model, New Hampshire will be able to clearly identify the costs and benefits of its spending on specific policies and prioritize the most successful and best-defined policies.



APPENDIX. WASHINGTON STATE³⁸

Program Evaluation

Since the 1990s, Washington State has used a cost-benefit analysis model in the criminal justice area to help identify cost-effective programs that consistently produce desired outcomes.³⁹ The use of such tactics have allowed Washington to decrease crime, juvenile arrest, and incarceration rates at a significantly greater level compared to the national average, saving \$1.3 billion per two-year budget cycle.⁴⁰ The Pew-MacArthur Foundation is helping other states to mimic Washington State's model in their own criminal justice policy, as well as aiding in the development of this model across other policy areas such as childhood education.

The Washington State Institute for Public Policy (WSIPP), a nonpartisan entity created by the state legislature in the early 1980s, conducts model analysis. WSIPP consists of a board that includes equal numbers of legislators and staff from both major parties, two appointees from the governor, and high-level staff from four universities in the state. The Washington legislature assigns WSIPP policy topics and it conducts studies using its own policy analysts, economists, specialists from universities, and consultants. It works closely with members and staff of the legislature, state agency staff, and experts in the field to provide objective and fact-driven analysis. WSIPP is funded primarily through Evergreen State College, which was given funds through an appropriations bill by the Legislature in 1983. Other projects deemed beyond the scope of funding provided to Evergreen State College may be funded by written legislation and/or appropriations bills.

The WSIPP Results First Model consists of eight steps of implementation: Analyze, Predict, Calculate, Assess, Rank, Identify, Assess, and Work. Below is a summary of the model and each of its steps.

1. Analyze all available research to systematically identify which programs work and which do not.
 - a. WSIPP analyzes studies that assessed the outcomes of related programs and policy options. They create what amounts to a meta-analysis, reviewing hundreds of studies on a wide range of programs. To do this, they first create an inventory of programs labeled as "Evidence-Based", "Researched-Based", or "Promising Practice."
2. Predict the impact of policy options.
 - a. An estimate of potential outcomes is created by applying combined evidence of all sufficiently rigorous studies to the state's own data. For example, WSIPP will research all studies available on early childhood education to predict a program's success in achieving key outcomes like reduction of child abuse, improvement in academic success, and reduction of substance abuse in families.
3. Calculate the potential return on investment (ROI) of policy options.



- a. WSIPP calculates the future cost for the state to produce the desired and predicted outcomes, and what the dollar value of these outcomes may be, in terms of savings and other benefits. Presents information in terms of Net Present Value (NPV), Cost-Benefit ratios, and ROIs. They then separate benefit projections into categories for program participants, taxpayers, and nonparticipants, which are subsequently combined to produce a total state bottom line.
- b. Example: Benefits generated by certain early childhood education plans are calculated, including higher earnings received by participants who are more likely to graduate from school, lower government criminal justice costs realized by participants who are less likely to commit crimes when older, and reduced costs to nonparticipants who are less likely to be crime victims.
4. Assess the investment risk if the initial assumptions behind the estimates turn out differently than predicted.
 - a. Statistical methods such as Monte Carlo simulations test key assumptions and measure the sensitivity of resulting outcomes. This determines the probability that a particular policy option would still produce net benefits even if the outcomes were different than initially predicted.
5. Rank the projected benefits, costs, and risks of all programs in a guide to policy options.
6. Identify ineffective programs that could be targeted for cuts or eliminations.
 - a. Based on the ranked list, the legislature cuts programs that are ineffective or not net-beneficial. This allows the legislature to focus on promoting only those programs deemed cost-effective or evidence-based rather than enacting across the board cuts.
7. Assess the benefits and costs of an interrelated package or “portfolio” of policies.
 - a. WSIPP will put together policy portfolios – a package of various policies and how they interact with each other, similar to an investment portfolio. In this way, the legislature can mix and match program outcomes to see which results in the highest ROI. Additionally, this allows the legislature to look at the policy area as a whole, and analyze how programs will affect the greater issue at hand, rather than focusing on the effects of a single program at a time.
 - b. In 2007, Washington acted on a portfolio of programs put together and suggested by WSIPP. With an investment of \$48 million, the state saved \$250 million by cancelling plans to build a new prison and enacting other programs instead.
8. Work with legislators and the executive branch to make these analyses highly accessible for policy and budget decision makers.
 - a. The analysis conducted by the WSIPP must be clearly communicated to state legislators in order for it to be effective. Data and reports are presented frequently, in terms that can be understood by someone without an advanced



The Nelson A. Rockefeller Center at Dartmouth College

The Center for Public Policy and the Social Sciences

degree in statistics, and researchers work closely with the executive branch throughout the entire process.



REFERENCES

- ¹ Muhlhausen, David B. *Evidence-Based Policymaking: A Primer*. The Heritage Foundation, 2015. <http://www.heritage.org/research/reports/2015/10/evidence-based-policymaking-a-primer>
- ² Baron, Jon and Ron Haskins. *The Obama Administration's Evidence-Based Social Policy Initiatives: An Overview*. The Brookings Institution, 2011. <http://www.brookings.edu/research/articles/2011/04/obama-social-policy-haskins>
- ³ *Evidence-Based Policymaking: A guide for effective government*. Pew-MacArthur Results First Initiative, 2014. <http://www.pewtrusts.org/~media/assets/2014/11/evidencebasedpolicymakingguideforeffectivegovernment.pdf>
- ⁴ Leonhardt, David. "The Quiet Movement to Make Government Fail Less Often." The New York Times, 2014. <http://www.nytimes.com/2014/07/15/upshot/the-quiet-movement-to-make-government-fail-less-often.html>
- ⁵ Muhlhausen, 2015.
- ⁶ Ibid.
- ⁷ Baron and Haskins, 2011.
- ⁸ Kalil, Tom. *Funding What Works: The Importance of Low-cost Randomized Controlled Trials*. The White House, 2014. <https://www.whitehouse.gov/blog/2014/07/09/funding-what-works-importance-low-cost-randomized-controlled-trials>
- ⁹ Leonhardt, 2014.
- ¹⁰ Rudd, Timothy, Elisa Nicoletti, Kristin Misner, and Janae Bonsu. *Financing Promising Evidence-Based Programs: Early Lessons from the New York City Social Impact Bond*. MDRC, 2013. <http://www.mdrc.org/publication/financing-promising-evidence-based-programs>
- ¹¹ Muhlhausen, 2015.
- ¹² *Inventory of Evidence-Based, Research-Based, and Promising Practices*. Washington State Institute for Public Policy, 2012. http://www.wsipp.wa.gov/ReportFile/1333/Wsipp_Inventory-of-Evidence-Based-Research-Based-and-Promising-Practices_Inventory.pdf
- ¹³ Muhlhausen, 2015.
- ¹⁴ Sec. 1208 (6). <http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- ¹⁵ Effective Public Health Practice Project. (1998). Dictionary for the Effective Public Health Practice Project Quality Assessment Tool for Quantitative Studies. Retrieved October 2008 from above web address. Corresponding article of interest: Thomas, B.H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 1(3), 176-184.
- ¹⁶ *Quantitative Research Assessment Tool*. CCEERC. http://www.researchconnections.org/childcare/datamethods/downloads/quantitative_research.pdf
- ¹⁷ *Report of the Governor's Commission on Innovation, Efficiency, and Transparency: Delivering 21st Century Operating Performance*. January, 2015.



<http://governor.nh.gov/commissions-task-forces/innovation/documents/iet-01-28-2015-final-report-plan.pdf>

¹⁸ Beauregard et al., 2015

¹⁹ *Delivering 21st Century Operating Performance*, 2015.

²⁰ *Evidence-Based Policymaking*. Pew-MacArthur.

²¹ Ibid.

²² Ibid.

²³ Ibid.

²⁴ *Results First in Your State*. Pew-MacArthur. Results First Initiative, 2014.

<http://www.pewtrusts.org/~media/assets/2013/results-first-in-your-state-brief.pdf?la=en>

²⁵ *Results-Based Management Handbook*. United Nations Development Group, 2011.

http://www.ilo.org/public/english/bureau/program/dwcp/download/undg_rbm1011.pdf

²⁶ *Federal Evidence-Based Innovation Programs*. Results for America, 2015.

<http://results4america.org/wp-content/uploads/2015/10/Innovation-fact-sheet.pdf>

²⁷ Baugus, Brian and Feler Bose. *Sunset Legislation in the States: Balancing the Legislature and the Executive*. Mercatus Center, 2015.

<http://mercatus.org/sites/default/files/Baugus-Sunset-Legislation.pdf>

²⁸ “Summary of State Legislation.” *Book of the States 2010*. Council of State Governments, 2010. http://knowledgecenter.csg.org/kc/system/files/Table_3.27.pdf, N.H. H.B. 458, 2011 Session. <http://www.gencourt.state.nh.us/legislation/2011/HB0458.html>

²⁹ Baugus and Bose, 2015.

³⁰ *Occupational Licensing: A Framework for Policymakers*. The White House, July 2015. https://www.whitehouse.gov/sites/default/files/docs/licensing_report_final_nonembargo.pdf

³¹ *Sunset in Texas: 2015-2017*. Texas Sunset Advisory Commission, 2015.

https://www.sunset.texas.gov/public/uploads/files/reports/Sunset%20in%20Texas_0.pdf

³² TX H.B. 1, 84th Legislature. (2015).

<http://www.capitol.state.tx.us/tlodocs/84R/billtext/pdf/HB00001I.pdf#navpanes=0>

³³ N.H. Const. Article 8. <https://www.nh.gov/constitution/billofrights.html>

³⁴ *Lean: Continuous Improvement*. New Hampshire Department of Administrative Services. <http://lean.nh.gov/>

³⁵ *Delivering 21st Century Operating Performance*, 2015.

³⁶ Ibid.

³⁷ *Evidence-Based Policymaking Commission Act of 2015*. H.R. 114, 114th Cong. (2015).

³⁸ *Better Results, Lower Costs: Washington State’s Cutting-Edge Policy Analysis Model*. Pew-MacArthur. Results First Initiative, 2012.

http://www.pewtrusts.org/~media/legacy/uploadedfiles/pca_assets/2012/resultsfirstwashingtoncasestudy.pdf

³⁹ Ibid.

⁴⁰ *Results First Clearinghouse Database*. Pew-MacArthur. Results First Initiative, 2015.

<http://www.pewtrusts.org/en/research-and-analysis/issue-briefs/2014/09/results-first-clearinghouse-database>