MEMORANDUM

JOINT BUDGET COMMITTEE

To	Joint Budget Committee Members
From	Christina Beisel, Craig Harper, and Vance Roper
Date	November 27, 2018
Subject	Internal JBC Evidence-Based Policy

EXECUTIVE SUMMARY

The Joint Budget Committee (Committee) has heard increasing discussion of Evidence-Based Policy (EBP) in recent years in two contexts: (1) state agencies have justified a variety of budget requests as "evidence based," prompting discussions of what that means; and (2) external groups and the Governor's Office have both emphasized a goal of increasing the use of EBP in Colorado. However, it has not always been clear what, exactly, "evidence based policy" meant in the context of specific programs or proposals.

Based on the Committee's interest, during the 2017 interim an internal JBC staff group formed an Evidence-Based Policy Team to conduct a research project to better understand EBP. Specific areas of investigation included: (1) background on EBP, including the meaning of EBP in the field, what constitutes "good" evidence, and how EBP is implemented; (2) the current use of EBP in Colorado, including both statutory requirements and executive branch actions; (3) other states' use of EBP, including legislative components encouraging or requiring use of EBP; (4) the limitations of EBP; and (5) potential paths forward should the General Assembly decide to expand the use of EBP in Colorado.

This process continued into the 2018 interim where JBC staff worked with multiple stakeholders, including the Colorado Evidence Based Policy Collaborative, to develop an internal policy proposal for the Committee's consideration. The Evidence-Based Policy Team has completed the project and submits the following recommendation to the Committee for discussion.

STAFF RECOMMENDATION

In order to facilitate the Committee's discussion of evidence based budget requests, staff recommends the Committee approve the following internal policies (discussed in greater detail later in the document):

1   Standard definitions that will be used to describe terms and processes for EBP programs. These definitions include the "Evidence-Based Continuum," which defines EBP tiers, as well as common terms used in evidence-based evaluations.

2   A process for highlighting and describing EBP programs in staff documents presented to the Committee.

EVIDENCE CONTINUUM

Proponents of EBP often discuss tiers of evidence and categorize programs based on the rigor of available evidence. These tiers range from a theory of change for new programs that have not yet been studied to a top tier of randomized controlled trials and quasi-experimental design including a systematic study and a control group for comparison. The tiers allow policymakers to categorize

programs, practices, or interventions based on the rigor of evidence supporting the practice. The Evidence-Based Policy Team recommends the Committee approve four tiers of evidence, as reflected in the Evidence-Based Continuum table below.

EVIDENCE CONTINUUM	EXAMPLES OF EVIDENCE	CONFIDENCE IN THE PROGRAM
Proven	• Two High Quality RCTs	High
Evidence-Informed	• 1 High Quality RCT • 2 High Quality QEDs	Moderate
Theory-Informed	• No Control or Comparison Groups	Moderate to Low
Opinion Based	• Satisfaction Surveys • Personal Experience • Testimonials	Low
Evaluation Investment	• No Existing Evidence • Quality Evaluation Planned	Unknown

DEFINITIONS USED IN EVIDENCE BASED EVALUATIONS

The Evidence-Based Policy Team recommends the following definitions for evaluation of evidence-based programs:

Evidence: Research and evaluations that indicate whether a program is capable of influencing and/or changing an outcome of interest.

Evidence Continuum: Evidence is built over time using a series of different research designs. The Evidence Continuum is the process of moving between the categories in the table above.

Comparison Group: A group (typically people) in an evaluation that either did not receive a program or were not randomly assigned to receive a program. The two groups are compared to measure a program's ability to influence and/or change an outcome of interest. Comparison groups are typically used in Quasi-Experimental Designs.

Control Group: A group (typically people) in an evaluation that have been randomly assigned to not receive a program. Data on the control group are compared to those receiving the program to measure a program's ability to influence and/or change an outcome of interest. Control groups are typically used in RCTs.

Outcome of Interest: The outcome that a program aims to influence and/or change. Program outcomes typically reflect behaviors, such as reducing recidivism or increasing academic achievement.

Quasi-Experimental Designs (QEDs): A research method that uses a comparison group. QEDs can produce high-quality evidence; however, they are not as reliable as RCTs in accounting for differences between subjects who receive a program and those who do not. Importantly, QED methods vary widely in their rigor, particularly in their ability to ensure program and comparison groups are equivalent on both observable and unobservable characteristics at the start of the program. However, some QEDs are highly capable of controlling for threats to internal validity and establishing causation.

Randomized Controlled Trial (RCT): A research method that uses a randomized control group, meaning that subjects are randomly assigned to either (i) a group that receives a program or (ii) a control group

that does not. Random assignment provides greater confidence that there are no systematic differences between the two groups. As a result, any difference in outcomes between the groups after the program can confidently be attributed to the program.

INTERNAL PROCESS

Evaluating an evidence-based program is time consuming and requires specific skills sets. The JBC staff does not have the capacity or the authority to perform full program evaluations including experimental designs, etc. However, the State has invested resources in evidence-based policy in recent years. For example, the Executive Branch has a dedicated EBP team that works with Departments to develop EBP programs and requests, which are highlighted in the annual budget submittal. Based on programs identified in the annual budget request, JBC staff can do a partial analysis on the program and how it fits into evidence based theory. For context, we expect to see fewer than 10 EBP requests across all departments in FY 2019-20.

In order to facilitate the Committee's discussions of evidence based budget requests, the Evidence-Based Policy Team recommends that the Committee approve the following internal policy on evidence-based evaluation. If a decision item, or any other programmatic request, is identified as an evidence-based program/request, JBC staff will:

1   Review the item to determine where the program falls on the evidence-based continuum.
2   Add a section in the staff briefing (if the item is a briefing issue) and figure setting documents titled "Evidence-Based Evaluation."
3   In the Evidenced-Based Evaluation section, JBC Staff will:
    a. Describe the tier of the continuum where the program falls, the confidence level in the program, and include a brief explanation on why the program falls into this area.
    b. Discuss the implementation plan for the program, or the lack of an implementation plan.
    c. Discuss the expected outcomes from the program's intervention.
4   Create an RFI that requires the Department or Agency running the program to report back on:
    a. The implementation process used for the program and a discussion on the fidelity of implementation for the program;
    b. Results in relation to the outcomes expected from the program;
    c. Lessons learned through implementation and administration of the program;
    d. Changes made based on the lessons learned; and
    e. Adjustments to outcomes based on lessons learned.
5   Report and discuss the information provided in the RFI to the Committee during the following fiscal year budget briefing process.

APPENDIX A COLLABORATIVE DOCUMENT

TO   **Vance Roper, Senior Legislative Analyst, Colorado General Assembly**

CC   **Craig Harper, Chief Legislative Analyst, Colorado General Assembly**

FROM   **The Colorado Evidence-Based Policy Collaborative**

DATE   **10/19/2018**

SUBJECT   **Evidence Standards**

Dear Vance,

The Colorado Evidence-Based Policy Collaborative thanks you for the opportunity to provide feedback
on the evidence standards you are recommending to the Joint Budget
Committee. Per your request, we reviewed standards for the following
terms:

1. Randomized Control Trial (RCT)
2. Quasi-Experimental Design (QED)
3. Control Group
4. Promising Practices

A glossary of terms is listed at the end of this document for reference.

The Collaborative's recommendations are that the Committee:

- References the Evidence Continuum listed in Table 1 (and illustrated in Figure 1);
- Focuses on the amount of *confidence* that different types of evidence provide (see Table 1) in terms of whether a particular outcome was caused by a given program;
- Understands that the more rigorous an evaluation, the more confidence we have in findings that demonstrate

Evidence Continuum	Examples of Types of Evidence	Confidence of Effectiveness, Ineffectiveness, or Harmfulness
Proven	• 2 high-quality RCTs	High
Promising	• 1 high-quality RCT, or • 2 high-quality QEDs	Moderate
Theory-Informed	• Evaluations with no control or comparison group	Moderate-Low
Opinion-Based	• Satisfaction surveys • Personal experience(s) • Testimonials	Low-None

whether our investments in programs achieve important outcomes for Coloradans _and_ do no harm;

- Recognizes that even programs with "Proven" evidence might not produce positive results if sufficient implementation resources are not invested; and
- Engages The Colorado Evidence-Based Policy Collaborative as a resource to vet and offer feedback on research, evaluation, and implemenation.

Table 1: The Evidence Continuum

The Evidence Continuum (Table 1) applies mostly to evaluations of individual service delivery (e.g. criminal and juvenile justice programs, behavioral health programs, child welfare programs, etc.). It does not best represent the highest level of evidence available –or even feasible– for many population-based strategies, including several implemented in state agencies such as public health (e.g., air quality), transportation (e.g., road safety), and agriculture (e.g., conservation).[1] The top evidence category listed in Table 1 is typically applied to programs that serve or engage individuals directly. Additionally, some programs (e.g., entitlements) are statutorily prohibited from randomizing their populations. This means that not all state programs can reach the level of "Proven." In these circumstances, certain QED study designs are recommended to evaluate these strategies. Since there are a variety of QED designs, each with different strengths, weaknesses and applications, their rigor should be considered in context. In addition to rigor, particular attention should also be paid to the quantity of QED studies evaluating these strategies and the consistency of their findings (e.g., one can be more confident in the evidence supporting a strategy that has been consistently shown to have important impacts across multiple high-quality QED studies, than one evaluated in a single QED study or one for which the evidence is not consistent across multiple studies). To this point, **the "best available evidence" is a principle that should be interpreted in context and determined by evaluation experts**.

It is also important to note that while most state programs will not fall into the "Proven" category, they still have value. We encourage the Committee to invest in moving programs along the evidence continuum (illustrated in Figure 1), or investing in the most appropriate scientific study design, so that the state's programs can build evidence to support their efforts or engage in process improvement efforts to achieve better outcomes.

The Evidence Continuum in Detail

Evidence building is an iterative process and starts once a program becomes "Theory-Informed" (see Figure 1). The best way to aid programs in building evidence, or in moving along the evidence continuum, is to fund evaluations of programs and support proper implementation to ensure fidelity to the chosen program/model. ***Implementation support is critical*** –especially for "Proven" programs, ***because proper training, materials, and funding (etc.) ensures programs are implemented as intended and therefore more likely to produce positive findings***.[2] Figure 1 provides guidance on the steps, or research activities, involved in building an evidence base, especially for programs that serve individuals and/or have a standardized curriculum. The figure also highlights how implemenation support takes place throughout the entire evidence-building process. As mentioned above, ***the further a program is along the continuum (as it moves towards "Proven"/Step 5), the more credible the findings are of program effectiveness, ineffectiveness, or harmfulness, and the more confidence that can be placed in the findings.***

<u>In Summary</u>
This memo focuses on: 1) identifying how confident we can be that a program is effective; 2) increasing
our confidence in the findings of a program's effectiveness through building an evidence base of more

---

[1] For example, the Colorado Department of Public Health and Environment utilizes several high-quality and respected resources including the Cochrane Review, the CDC Community Guide, the World Health Organization, the National Academies of Sciences, Engineering, and Medicine (previously the Institute of Medicine), and other reputable, peer-reviewed research to identify recommended strategies with substantial evidence of impact.

[2] A good example of this is an implementation of the Functional Family Therapy program on the west coast. Through evaluation it was discovered that gaps in the program's

rigorously designed evaluations; and 3) underscoring the important concept that evaluations produce findings that range from effective to harmful, and that confidence in those findings is associated with the level of rigor in the research study reporting outcome results.

The Colorado Evidence-Based Policy Collaborative would like to thank the Committee for considering our recommendations and invites the Committee to reach out with questions.

Glossary of Terms

Figure 1: The Evidence Continuum
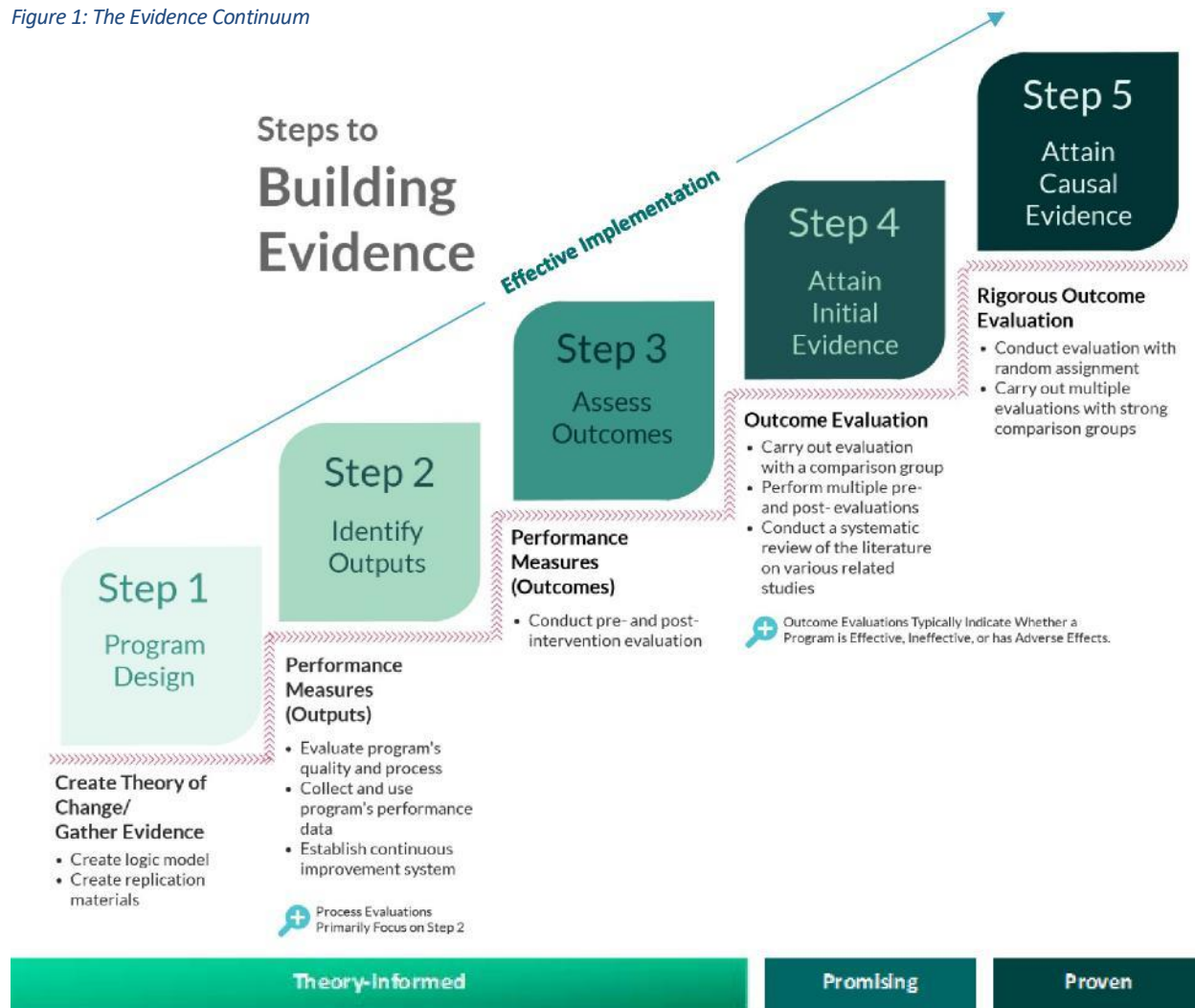
Evidence: Research that indicates whether a program is capable of influencing and/or changing an outcome of interest.

Evidence Continuum: Evidence is built over time using a series of different research designs. The graphic "Steps to Building Evidence" (Figure 1) displays this continuum.

Comparison Group: A group (typically people) in an evaluation that either did not, or were not *randomly* assigned to, receive a program. Data on the *comparison group* are compared to how other tested subjects (those receiving the program) do to benchmark and measure a program's ability to influence and/or change an outcome of interest. Comparison groups are typically used in QEDs.

Control Group: A group (typically people) in an evaluation that have been *randomly* assigned to <u>not</u> receive a program. Data on the *control group* are compared to how other tested subjects (those receiving the program) do to benchmark and measure a program's ability to influence and/or change an outcome of interest. Control groups are typically used in RCTs.

Outcome of Interest: The outcome that a program aims to influence and/or change. Program outcomes typically reflect behaviors, such as reducing recidivism or increasing academic achievement.

Quasi-Experimental Designs (QEDs): A research method that uses a comparison group. QEDs can produce high-quality evidence; however, they are generally not as reliable as RCTs in accounting for differences between subjects (typically people) who receive a program and those who do not. Importantly, QED methods vary widely in their rigor, particularly in their ability to ensure program and comparison groups are equivalent on both observable and unobservable characteristics at the start of the program. However, some QEDs are highly capable of controlling for threats to internal validity and establishing causation (like a well-done RCT), such as a well-designed and executed instrumental variable analysis, regression discontinuity design or comparative interrupted time series design. Experienced researchers determine which designs are most suitable and/or appropriate for the program in consideration of its setting.

Randomized Controlled Trial (RCT): A research method that uses a randomized control

group, meaning that subjects (typically people) are randomly assigned to either (i) a group that receives a program or (ii) a control group that does not. Random assignment ensures to a high degree of confidence that there are no systematic differences between the program and control groups in their observable or unobservable characteristics at the start of the program. As a result, any difference in outcomes between the groups after the program can confidently be attributed to the program. It should be noted that with this evaluation design, either group can concurrently access any other available programs as needed.

Colorado Evidence-Based Policy Collaborative Members

- Adrienne Russman, Director of Collective Impact, Uncharted
- Ali Maffey, Policy and Communication Unit Supervisor, Colorado Department of Public Health and Environment
- Ann Renaud Avila, ARA Strategies
- Bill Woodward, Director, Training and Technical Assistance, Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado
- Brian Bumbarger, PhD, Adjunct Research Associate, Colorado State University Prevention Research Center; Adjunct Research Fellow, Griffith University Institute of Criminology; Consultant, Annie E. Casey Foundation
- Cindy Eby, Founder/CEO, ResultsLab
- David Anderson, Director of Evidence-Based Policy, Laura and John Arnold Foundation
- Diane Pasini-Hill, EPIC Manager, Division of Criminal Justice, Department of Public Safety
- Jessica Corvinus, Research and Evidence-Based Policy Manager, Colorado Governor's Office
- Kristen Pendergrass, Principal Associate, Pew–MacArthur Results First Initiative
- Kristin Klopfenstein, PhD, Director, Colorado Evaluation and Action Lab
- Lisa Hill, Executive Director, Invest In Kids
- Pamela Buckley, PhD, Director of Blueprints for Healthy Youth Development
- Sarah Prendergast, Doctoral Student, Colorado State University
- Tiffany Madrid, Research and Evidence-Based Policy Lead Analyst, Colorado Governor's Office
- Tiffany Sewell, Collaborative Management Program Administrator, Colorado Department of Human Services