Jurnal Filsafat, ISSN: 0853-1870 (print); 2528-6811(online)
Vol. 35, No. 1 (2025), p. 125 – 153, doi: 10.22146/jf.103353

"TEST, LEARN, AND LISTEN": RETHINKING THE EPISTEMOLOGICAL ASSUMPTION OF EVIDENCE-BASED POLICYMAKING

Taufiqurrahman
Faculty of Philosophy, Universitas Gadjah Mada

Arga Pribadi Imawan
Department of Politics and Government, Faculty of Social and Political Sciences, Universitas Gadjah Mada

Agus Wahyudi
Faculty of Philosophy, Universitas Gadjah Mada
Email: awahyudi@ugm.ac.id

Abstrak

Pembuatan kebijakan berbasis-bukti (evidence-based policymaking/EBP) mengandalkan asumsi epistemologis bahwa bukti dari uji acak terkendali (randomized controlled trial/RCT) adalah bukti terbaik bagi formulasi kebijakan, sementara testimoni ahli adalah yang terburuk. Paper ini berargumen bahwa meski RCT adalah sumber bukti empiris yang berharga bagi intervensi kebijakan, mereka tidak cukup dengan sendirinya untuk mendukung formulasi kebijakan berbasis-bukti. Melalui lensa kerangka kerja kausasi INUS, kami mendemonstrasikan bahwa keefektifan sebuah kebijakan dipengaruhi oleh interaksi kompleks antara berbagai faktor kontekstual yang tidak dapat ditangkap oleh RCT semata. Karenanya, kami mengadvokasi integrasi pengetahuan kontekstual dan kualitatif, termasuk testimoni dari ahli dan anggota masyarakat, untuk melengkapi temuan RCT. Pengetahuan tambahan ini menyediakan wawasan mengenai dimensi sosial, kultural, dan subjektif dari populasi sasaran sehingga dapat mengakomodasi motivasi, preferensi, dan faktor lain yang dapat memengaruhi keberhasilan kebijakan secara signifikan. Dengan mengomparasikan perspektif reduksionis dan non-reduksionis mengenai guna testimoni dalam kebijakan berbasis-bukti, kami mengargumentasikan pendekatan seimbang yang menghargai testimoni kredibel sebagai hal

esensial untuk memahami konteks. Pada akhirnya, paper ini menggarisbawahi pentingnya pendekatan bukti beragam dalam membuat kebijakan publik yang efektif dan peka-konteks

***Kata kunci**: Kebijakan berbasis-bukti, uji acak terkendali, testimoni, epistemologi terapan.*

Abstract

Evidence-based policymaking (EBP) relies on an epistemological assumption that evidence from randomized controlled trials (RCTs) is the finest evidence for policy formulation, while expert testimony is the poorest one. This paper argues that while RCTs are a valuable source of empirical evidence for policy interventions, they are not sufficient on their own to support evidence-based policy formulation. Through the lens of the INUS framework of causation, we demonstrate that the effectiveness of a policy is influenced by a complex interplay of contextual factors, which RCTs alone cannot capture. Hence, we advocate for the integration of contextual and qualitative knowledge including testimonies from experts and community members, to supplement RCT findings. This additional knowledge provides insights into the social, cultural, and subjective dimensions of the target population, addressing motivations, preferences, and other factors that can significantly impact policy success. By comparing reductionist and non-reductionist perspectives on the use of testimony in evidence-based policy, we argue for a balanced approach that values credible testimonies as essential to understanding context. Ultimately, this paper underscores the importance of a multifaceted evidence approach in crafting effective, context-sensitive public policies.

***Keywords**: Evidence-based policy, randomized controlled trials, testimony, applied epistemology.*

Received: January 03, 2025| **Reviewed**: January 15, 2025| **Accepted**: February 28, 2025

INTRODUCTION

Over the last two decades, evidence-based policymaking has gained increasing popularity. The central tenet of this model is that public policymaking should be grounded in the best-available evidence (Pérez-González, 2024). This approach responds to a policymaking paradigm historically driven by the political ideologies of policymakers or expert opinions (Pearce & Raman, 2014). The evidence-based policymaking movement began to take shape in the late 1990s (Littell & White, 2018; Parsons, 2002), drawing inspiration from the earlier emergence of evidence-based medicine (Baron, 2018; Smith, 1996). Essentially, it is the application of evidence-based practice to the social realm.

Proponents of evidence-based policy (EBP) argue that it is superior to policy based on ideology or expert opinion because the interventions it formulates have been empirically tested to ensure they produce expected outcomes (Goldacre, 2011). This superiority is underscored by the evidence underpinning each policy intervention. In both medical and policy contexts, evidence-based practice assumes an epistemological hierarchy of evidence to guide intervention formulation. Various versions of evidence rankings exist, but they consistently place randomized controlled trials (RCTs) at the top and expert opinion at the bottom (Canadian Task Force on the Periodic Health Examination, 1979; Sackett, 1989; SIGN, 2011; Weightman et al., 2005). Consequently, evidence from RCTs is often referred to as the "gold standard" (Baron, 2018; Haynes et al., 2012).

A randomized controlled trial (RCT) is an empirical method used to test the effectiveness of a policy intervention or medical treatment. For instance, to test the efficacy of a headache reliever, pharmacologists might recruit 500 individuals with headaches and randomly assign them into two groups, constituting a study population. The first group receives the headache reliever being tested (intervention group), while the second group receives a placebo (control group). If the number of people whose headaches subside is significantly greater in the intervention group than in the

control group, it indicates the drug's effectiveness. Thus, the drug can be marketed to and consumed widely by individuals who suffer from headaches, representing the target population.

For many years, evidence-based practices have been relying on an epistemological assumption that RCT is the most reliable empirical method for testing the effectiveness of an intervention or the causal relationship between an intervention and a desired outcome. In other words, RCTs are highly valued for their ability to assess the internal validity of an intervention (Pérez-González, 2024). However, in the context of policymaking, RCTs are used not only to test the effectiveness of a policy intervention in the study population but also to justify extrapolating the intervention to the target population (external validity). Consequently, recent years have seen increasing criticism of RCT use in public policy formulation.

The primary criticism is that RCT results cannot be automatically extrapolated from one population to another, as effectiveness in one setting does not guarantee effectiveness in another (Cartwright, 2012; Cartwright & Hardie, 2012; Cowen et al., 2017; Deaton & Cartwright, 2018; Knox et al., 2018). This issue is exemplified by the California Class Size Reduction program. Since 1985, the State of Tennessee conducted an RCT to assess the efficacy of class-size reduction in improving students' academic performance. The trial indicated that reducing class size could substantially enhance student learning in both mathematics and reading (Mosteller, 1995). In 1996, partly inspired by these findings from Tennessee, California enacted legislation to cap class sizes in kindergartens and allocated $1 billion annually to support districts in implementing class size reduction. However, the intervention yielded minimal impact. The final program report concluded that the correlation between class size reduction and student achievement was inconclusive (Hayward & Kirst, 2002).

The criticism highlights a fundamental limitation of using RCTs in public policymaking, specifically that it overlooks the contextual and qualitative dimensions of policy formulation.

Therefore, this paper seeks to critically assess and rethink the epistemological assumption of EBP that favors evidence from RCTs, on the one hand, and disfavors testimony, on the other hand. In other words, what kinds of evidence should policy-makers adopt to formulate an intervention? While recent literature has critiqued this assumption, much of the focus has been on problematizing RCTs. Some studies have shown that RCTs face significant challenges in evaluating the effectiveness of relevant policies and interventions (Gamoran, 2018; Gelman, 2018; Kemm, 2006; Khosrowi & Reiss, 2019; La Caze & Colyvan, 2017; Montuschi, 2009). Additionally, other research has demonstrated that RCTs often fall short in supporting the extrapolation or generalization of causal claims beyond the populations studied (Cartwright, 2012; Cartwright & Hardie, 2012; Deaton & Cartwright, 2018; Knox et al., 2018). Furthermore, ethical concerns have been raised regarding RCTs, as they may fail to identify potential side effects (La Caze & Colyvan, 2017) and often lack the inclusion of informed consent and equipoise in their trials (Lisciandra, 2020). Beyond demonstrating the insufficiency of RCTs, this paper also argues that expert and local testimonies can improve the shortcomings of RCTs.

Accordingly, this paper is divided into five sections. The first section discusses the operation of RCTs within evidence-based policymaking. The second section presents a theoretical framework for assessing the effectiveness of policy interventions, providing a case against the sufficiency of RCTs in EBP. The third section urges to complement RCTs with contextual and qualitative knowledge of the target population. The final section provides a justification of why testimony can be a valuable supplement to evidence from RCTs.

RCTs in Public Policymaking

In modern states, the government plays a central role due to its obligation and authority to make decisions affecting citizens' lives. Government decisions that have an influence on the life of citizens are known as public policies (Peters, 2010). In essence,

public policy encompasses whatever governments choose to do or not do (Dye, 2013). Thomas A. Birkland defines policy more specifically as:

> "A statement by the government of what it intends to do, such as a law, regulation, ruling, decision, order, or a combination of these. The lack of such statements may also be an implicit statement of a policy not to do something," (Birkland, 2020, p. 6).

Given its impact on public life, every policy must undergo at least five stages. First, identify the problems that need urgent attention (agenda setting). Second, formulate policies available to address the identified problems. Third, decide on the policies to be implemented. Fourth, implement the decided policies. Finally, evaluate the implemented policies (Howlett & Giest, 2015). The use of RCTs in public policymaking can occur in the second or fifth stages, depending on whether the policy being tested is widely implemented or new.

Regardless of their position in the policymaking process, RCTs aim to test the effectiveness of policy interventions. Haynes et al. (2012) from the Behavioral Insights Team outline nine steps for implementing RCTs in a public policy context. The first step involves identifying two or more policies to compare. Notably, one of the policies can be a passive intervention, where no additional treatment is applied beyond standard care. For instance, the effectiveness of mask-wearing in reducing the spread of COVID-19 can be evaluated by comparing a group wearing masks with a control group receiving no intervention.

The second step is to determine the expected outcomes from implementing a policy intervention. For example, the Indonesian government has implemented a 12-year compulsory education policy with the expectation that this policy will increase the school participation rate of Indonesian children at the high school level. Therefore, the school enrollment rate will be calculated to measure the effectiveness of the policy.

The third step is to specify the randomization unit or entity that will be randomly assigned to the intervention group and the control group. The unit of randomization can be an individual, group of individuals, institution, or region. To assess the effectiveness of the 12-year compulsory education program, for instance, the region can be used as the randomization unit. Once the randomization unit is determined, the fourth step is to decide how many units need to be tested to obtain robust results.

The fifth step is the core component of RCTs, which involves carrying out randomization. All units to be tested are randomly assigned to either an intervention group or a control group. The intervention group will receive the policy intervention, while the control group will not. Random distribution is crucial to avoid bias and to ensure that both the intervention and control groups have comparable characteristics. This ensures that any observed changes in the intervention group after implementing the policy can be attributed to the intervention itself, rather than other factors. This aspect is what makes RCTs superior to other methods.

After randomization is carried out, the sixth step is to apply the policy intervention to the group designated as the intervention group. This implementation should mirror how the policy would be executed on a wider scale if it proves effective according to the RCT. This approach ensures consistency between the trial implementation and the wider application of the policy. In this way, we can expect that the outcomes observed during the trial will be replicated when the policy is implemented more broadly.

The seventh step is to measure and determine the impact of the policy interventions that have been implemented. The timeframe for assessing the results of a policy varies and can range from a week, a month, a year, to five years after the policy's initial implementation. For instance, in the case of the compulsory education program, the results may only become evident after a minimum of three years. Conversely, for regulations requiring the use of masks to reduce the risk of transmitting COVID-19, results can be observed within weeks or even days.

Upon obtaining the results, the eighth step involves adapting these findings into policy. If the results are favorable, the tested policy intervention may continue to be implemented and potentially expanded in scope. Conversely, if the results are unfavorable or neutral, the policy intervention should be discontinued to ensure the appropriate allocation of resources and to avoid adverse effects. After incorporating the results of randomized controlled trials (RCTs) into policy, it is essential to revisit the initial step to re-evaluate or refine policies that are already in widespread use.

Haynes et al. (2012) consolidate the nine steps into three overarching phases: “test, learn, and adapt”. The "test" phase, encompassing steps one through six, involves evaluating the effectiveness of a policy intervention through robust and reliable methods. The "learn" phase, which includes the seventh step, focuses on analyzing the test results to assess whether the intervention effectively achieves the desired outcomes. The "adapt" phase, incorporating steps eight and nine, entails modifying the policy based on the test results.

Proponents of RCTs provide several reasons for their essential role in public policy-making. Buck & McGee (2015) argue that RCTs represent "the best way to learn whether a social program or policy works as intended ... because they make it possible to isolate the effect of a program from complicating factors, even those that are unseen." Haynes et al. (2012) present various case studies demonstrating the superiority of RCTs. They highlight interventions traditionally believed to be beneficial that RCTs have shown to be ineffective or even detrimental. For instance, RCTs have challenged the efficacy of steroid treatments for head injuries (Edwards et al., 2005) and the Scared Straight Program, which aims to deter youth from criminal behavior by exposing them to the harsh realities of prison life (Petrosino et al., 2013).

Certain programs that seem intuitively effective at achieving desired outcomes, such as offering incentives to encourage adult learners to attend literacy classes, may not always perform as expected. RCTs revealed that individuals receiving incentives

attended fewer classes compared to those who did not receive incentives (Brooks et al., 2008). Conversely, RCTs can also validate the effectiveness of interventions that initially appear less promising. For example, sending text messages to individuals who have failed to pay court fines has been shown to significantly increase payment rates (Haynes et al., 2012). This approach has since been adopted by various organizations to enhance payment compliance and reduce instances of late payments. For these reasons, the use of RCTs in policy-making continues to be strongly advocated and promoted.

RCTs and the Theoretical Frameworks of Causation

RCTs are a methodological approach designed to rigorously evaluate the efficacy of policy interventions in achieving predetermined outcomes. Therefore, the principal aim of RCTs is to ascertain a causal relationship between the intervention and the targeted outcome. This is accomplished by the random allocation of subjects into either intervention or control groups, thereby mitigating potential confounding variables. If the intervention group successfully achieves the targeted outcome while the control group does not, it can be inferred that a causal relationship exists between the intervention and its outcome.

Such experimental designs presuppose a specific theoretical framework concerning causation, particularly the counterfactual theory of causation. According to this theory, x causes y if and only if, if x had not occurred, y would not have occurred (Menzies & Beebee, 2024). In other words, the causal relationship is based on counterfactual dependence: whether x can be considered the cause of y depends on what would have transpired had x not occurred. If x did not occur and y still transpired, then x was not the cause of y.

This counterfactual analysis of causation was proposed by David Lewis:

> "We think of a cause as something that makes a difference, and the difference it makes must be a difference from what would

> have happened without it. Had it been absent, its effects—some of them, at least, and usually all—would have been absent as well." (Lewis, 1973, p. 557).

Lewis proposed a counterfactual analysis of causation, drawing inspiration from David Hume, who is more commonly associated with the regularity analysis of causation. In An Enquiry Concerning Human Understanding, Hume defines 'cause' as follows:

> "an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or, in other words, where, if the first object had not been, the second never had existed," (Hume, 2007, p. 56).

For Lewis, what Hume wrote in the second sentence is not a restatement of the regularity analysis given in the previous sentence. Instead, Hume is indeed providing two types of analysis of causation: regularity analysis and counterfactual analysis.

The main idea of counterfactual analysis is the assumption that if x did not occur, then y would also not occur. In the design of RCTs, this concept is operationalized through the inclusion of a control group that does not receive the intervention. If the control group, which is not exposed to the intervention x, produces the same outcome y as the intervention group, then x cannot be deemed the cause of y, as y could have occurred independently of x. Consequently, the presence of a control group is indispensable in RCTs, in addition to randomization.

However, there are several criticisms of the counterfactual analysis of causation underpinning RCTs. This analysis seeks to reduce causation to counterfactual dependence, positing that any factor deemed a cause must exhibit counterfactual dependence, and vice versa. In other words, counterfactual dependence is considered both necessary and sufficient for causation (Moore, 2009, p. 394). Several philosophers have challenged this assumption in various ways. First, they question the sufficiency of counterfactual

dependence as a criterion for causation, pointing to instances where counterfactual dependence is present without a causal relationship. Second, they challenge the necessity of counterfactual dependence for causation, identifying cases where a causal relationship exists despite the absence of counterfactual dependence.[1]

One significant limitation of counterfactual analysis is its inability to capture the complexity of causation. In the case of x does not cause y not in isolation; rather, x typically requires a set of prerequisites or background conditions to effectively produce y. For example, friction (x) causes a match to light (y), but friction alone is not sufficient for ignition. The match also needs to be dry, properly designed, and rubbed in an environment with available oxygen. If, for example, the match is rubbed in a wet condition or in the absence of oxygen, it will not ignite. Consequently, in a counterfactual analysis, factors such as oxygen and the match's dryness would need to be considered as part of the causal explanation for the match lighting. However, including these background conditions as causes seems implausible, as they merely facilitate the lighting process rather than directly cause it.

A similar issue arises in the context of public policy. A policy intervention can only achieve its desired outcomes if a number of prerequisite background conditions are met. The class size reduction initiatives in Tennessee and California provide a relevant example. While class size reduction proved effective in Tennessee, it failed to improve student achievement in California. This discrepancy occurred because the implementation in California lacked the necessary background conditions for effective intervention, such as sufficient classroom space and a sufficient number of qualified teachers to accommodate the increased number of classes resulting from the reduction in class size (Cartwright & Hardie, 2012, p. 65).

---

[1] For a more detailed exposition of objections to counterfactual analysis, please see Moore (2009).

Counterfactual analysis, when confronted with causal complexity, faces two potential pitfalls: either overlooking the background conditions necessary for the effectiveness of a policy intervention or erroneously attributing the background conditions as the primary causes. The first issue is illustrated by the class size reduction case. The second issue is exemplified by the policy of mandatory helmet use for cyclists intended to reduce accidents resulting in head injuries. Case-control studies indicate that cyclists who wear helmets experience fewer head injuries compared to those who do not (Dorsch et al., 1987; Thompson et al., 1989; Wasserman et al., 1988). Consequently, the British Medical Association (BMA) advocates for mandatory helmet use. However, a time-series study conducted in regions that have enacted helmet use regulations did not observe a decrease in head injuries among cyclists following the implementation of the regulation. In some instances, this study even observed an increase in head injuries among cyclists after the regulation was enforced (Robinson, 2006).

Intuitively, helmets, as head protection devices, are easily associated with a reduction in head injuries among cyclists. However, this perspective fails to account for other factors that are also causally relevant, such as the behavior of cyclists, motorcyclists, and car drivers, as well as road conditions and vehicle density. The use of a helmet may, in fact, lead to increased risk-taking by cyclists due to a false sense of security, or it might prompt motorcyclists and car drivers to be less cautious around cyclists, assuming that the helmets provide sufficient protection (Adams & Hillman, 2001). Therefore, despite wearing helmets, cyclists may still face a high risk of head injuries. Helmets are necessary but not sufficient on their own to reduce the incidence of head injuries. The counterfactual analysis of the case-control study erroneously treats helmet use as the primary factor in reducing head injuries, when it is merely a necessary background condition. The more crucial factor in minimizing head injuries among cyclists is the behavior of both cyclists and drivers, which should prioritize safety and orderliness.

To address the limitations of counterfactual analysis in capturing causal complexity, we need another causal model. Nancy Cartwright and her colleagues have proposed the INUS conditions model as a more robust approach for capturing the complexity of causal phenomena (Cartwright & Hardie, 2012, pp. 63–64; Cartwright & Stegenga, 2011, pp. 301–302). Originally formulated by Mackie (1965), the INUS model aims to better accommodate the complexity inherent in causal phenomena. For instance, consider a fire in a house: an electrical short circuit might be identified as a cause of the fire. However, an electrical short circuit is not a necessary condition for a fire; the house could also burn due to other factors, such as a gas stove malfunction or a cigarette butt igniting gasoline. Furthermore, an electrical short circuit alone is not sufficient to cause a fire; if there were no flammable objects near the short circuit, the fire would not have occurred.

Thus, the electrical short circuit is neither a necessary nor a sufficient condition for a fire to occur. The question then arises: why is it considered a cause of the fire? According to Mackie (1965, p. 245), "the short-circuit which is said to have caused the fire is thus an indispensable part of a complex sufficient (but not necessary) condition of the fire." This means that the fire was actually caused by a complex condition comprising an electrical short circuit, a flammable object near the short circuit, and the absence of fire prevention equipment. Each component of this complex condition is "insufficient" (I) but "necessary" (N) for producing a fire; in other words, the absence of any one component would prevent the fire from occurring, as all are required for ignition. When these components are present together, they form a condition that is "unnecessary" (U) but "sufficient" (S) for a fire to occur. The term "unnecessary" refers to the fact that the absence of this particular complex condition does not guarantee that a fire will not occur, as other complex conditions might also lead to a fire, such as cigarette butts on a gasoline-spilled floor with flammable objects nearby. Hence, an electrical short circuit is considered a cause of the fire

because it is an **i**nsufficient but **n**ecessary part of a complex condition that is **u**nnecessary but **s**ufficient for the fire to occur.

The INUS conditions causal model effectively explains the issues encountered with the policies of class size reduction and mandatory helmet use. In the case of class size reduction, small class sizes represent a necessary but insufficient condition for enhancing students' academic achievement. For this intervention to be effective, it must be complemented by competent teachers and adequate classroom space. Similarly, the use of helmets is not sufficient on its own to reduce the incidence of head injuries among cyclists. Effective reduction in head injuries also requires safe cycling and driving behavior, as well as favorable road conditions. Thus, both policies require additional conditions to achieve their intended outcomes, illustrating the applicability of the INUS model in addressing causal complexity.

The Significance of Contextual and Qualitative Knowledge in EBP

Evidence-based policymaking has increasingly focused on empirical rigor as the basis of successful public policies. The tendency to laud these quantitative methods, particularly randomized controlled trials, results in an inability to isolate causal relationships while ignoring the more general sociocultural, political, and environmental contexts within which these policies operate. This can lead to a policy mismatch between design and real-world conditions. This helps ensure that policies are evidence-based, practical, culturally sensitive, and capable of addressing the complexities of human behavior and societal structures with knowledge about the target population.

Contextual knowledge involves unique social, cultural, economic, and institutional conditions that sincerely serve to shape life in a community. These are necessary to understand the policy reception, implementation, and sustainability in those contexts. So often, the contextual factors determine intervention success or failure even when the policy being implemented is based on sound

empirical evidence. For example, what works in health policy in one regional setting may fail in another because of disparate infrastructure, cultural practices, or community trust in government institutions (Cartwright & Hardie, 2012).

A prominent example of this is the use of public health measures in combating the COVID-19 pandemic. Quantitative results from RCTs indicated that mandates to wear masks and maintain social distancing can help limit virus transmission. On the contrary, qualitative studies showed significant compliance barriers, including cultural resistance, misinformation, and logistical challenges in accessing masks, among other things (Haynes et al., 2012). Policymakers engaging with local communities and considering these contextual factors were better placed to design interventions that addressed not only the "what" but also the "how" of implementation and were better able to calibrate their strategies to the realities of their populations.

It also takes deep contextual knowledge to comprehend what ripple effects policies will have. The narrowness of an intervention itself can sometimes yield unintended consequences when the broader social dynamics are disregarded. In some societies, for example, policies enacted to increase economic participation by women run counter to traditional gender norms in those societies and fuel community backlash or resistance. In this situation, qualitative research can easily detect potential conflicts and propose avenues to reduce social friction, thus allowing smoother policy implementation (Montuschi, 2009).

Qualitative knowledge examines subjective elements of human life, such as motivations, values, preferences, and perceptions. While quantitative methods have strengths in measuring outcomes, they facilitate insight into mechanisms behind those outcomes and the issues affecting individual and collective behavior. Therefore, integrating qualitative data with statistical findings will better map policies into the concerns and aims of the communities being served.

For instance, quantitative data could suggest that smaller class sizes increase student achievement in education reforms. At the same time, qualitative research would uncover the actual reasons for such improvement, like more teacher-student interaction or increased student participation (Mosteller, 1995). If policymakers do not understand such qualitative dimensions, they are at risk of misinterpreting quantitative findings or missing those critical variables that will contribute to the success of a policy.

Similarly, during the vaccination process for COVID-19 in Indonesia, quantitative evidence covered the ability of vaccines to scale down morbidity and mortality rates. However, qualitative research proved crucial in handling the high level of widespread vaccine hesitancy. Working with religious leaders and issuing a halal certification for the vaccines showed the recognition of the long-held cultural and religious beliefs by the predominantly Muslim government of their people (BPMI Setwapres, 2021). This culturally sensitive approach dramatically boosted public confidence and vaccine uptake while underlining qualitative knowledge's function in successful policy intervention. Besides knowledge of individual and cultural factors, qualitative research is essential for targeting group power issues.

Policies often affect marginalized groups differently than they do more privileged populations. For instance, city housing programs end up ignoring and displacing low-income residents without such involvement during planning and design. Such hidden inequalities uncovered through ethnographic studies and participatory methods will help policymakers develop just and fair solutions (Montuschi, 2009).

1. Theoretical Foundations: A Case for Integration

Contextual and qualitative knowledge in EBP challenges the classical hierarchy of evidence, wherein RCTs are considered at the top and qualitative evidence at the bottom. Philosophers such as Nancy Cartwright argue that causation in social systems is

inherently complex and cannot be fully captured by controlled experiments alone (Cartwright, 2012). The INUS framework—Insufficient but Necessary parts of an Unnecessary but Sufficient condition—provides a theoretical basis for this argument. According to this framework, causation is rarely attributable to a single variable but is the product of interdependent conditions that collectively influence outcomes (Mackie, 1965).

For instance, the effectiveness of a policy aimed at reducing traffic-related accidents would involve not only the enforcement of wearing helmets but also the development of road infrastructure, control of the movement speed of vehicles, and encouragement of safe driving. Although RCTs may isolate what happens because of the helmets, they cannot capture these interrelated factors. Qualitative approaches can yield more detailed insight into the design of comprehensive intervention interviews among road users and observational studies.

Qualitative knowledge is extended further by these epistemological theories of testimony. This may take the form of testimonial evidence provided by experts or local community members that quantitative approaches could not capture. Although traditions in epistemology, like reductionism, argue that testimony has to be corroborated from other sources, non-reductionist thinking accepts testimony as an independent, legitimate source of knowledge in its own right (Leonard, 2023; O'Brien, 2024). These perspectives support the belief in qualitative insight into the most complex policymaking environments where statistical data simply cannot give a complete picture.

Lack of consideration for contextual and qualitative knowledge can be imbued with profound ethical and practical implications. Policies that do not consider the lived experience of their subject communities are ineffective or, worse still, destructive. In several cases, rural livelihood development programs have resulted in unplanned changes, such as the displacement of communities or environmental degradation, due to the local participants' lack of consideration at the design stage (Montuschi,

2009). Such failures, however, point to the moral commitment of policymakers to listen to and involve the communities they serve in policymaking processes.

Policies that render qualitative knowledge invalid might involuntarily sustain systemic injustices. The example of environmental policies that have brought about a shift to reduce carbon emissions by increasing public transportation shows such moves, while theoretically fitting for all citizens, may have potentially significant impacts in negative ways on the poorer sections of society if reasonable alternative transportation is unavailable. These will ensure that policy addresses varied stakeholders' multi-dimensional needs, raising equity and inclusivity.

2. Shifting Towards a Balanced Evidence Paradigm

The actual value of EBP will accrue when policymakers are in a position to embrace a balance in the evidence paradigm valued in quantitative and qualitative insights. Such a paradigm acknowledges that while RCTs provide fundamental data on cause-and-effect relationships, these must indeed be supplemented with contextual and qualitative data to ensure real-world applicability. A perfect example of such a balanced practice is the participatory budgeting model in many cities, such as Porto Alegre, Brazil.

The key difference from other, more traditional EBP models includes the iterative "test, learn, and listen" process. In essence, policymakers are called to test interventions in controlled settings and listen to voices from within the most affected communities, learning from their lived experiences to refine policy continuously. That is, evidence would be assuredly accurate but also relevant and actionable. To effectively incorporate contextual and qualitative knowledge into EBP, policymakers should: 1) C*onduct mixed-methods research*—i.e., integrating quantitative and qualitative approaches to capture both outcomes and the mechanisms driving those outcomes; 2) *Engage local communities*—i.e., use participatory

methods to gather insights from those directly affected by policies; 3) *Consult interdisciplinary experts*—i.e., collaborate with anthropologists, sociologists, and other specialists who can provide contextual knowledge; 4) *Prioritize inclusivity*—i.e., ensure that marginalized voices are represented in policymaking processes; and 5) *Adopt iterative policy design*—i.e., continuously refine policies based on feedback from qualitative evaluations and field observations.

Incorporating contextual and qualitative knowledge in EBP is not a supplement to quantitative methods such as RCTs but is part of effective policymaking. Because they understand the target populations' distinctive characteristics and lived experiences, policymakers are better placed to design interventions that work. As the state of California's Class Size Reduction program and the Indonesian vaccination initiative both illustrate, those policies that do not consider context are likely to be inefficient or even counterproductive. From this perspective, EBP needs to consider a holistic approach in which the core of empirical rigor values the richness of contextual qualitative insights.

Evidential Supplement from Testimony

We have demonstrated that evidence derived solely from randomized controlled trials (RCTs) is insufficient as a foundation for public policy. Such evidence must be augmented with contextual and qualitative insights specific to the target population. In this section, we will argue that the testimonies of experts and local citizens can effectively address the limitations inherent in RCTs. This discussion centers on two primary questions: (1) Can testimonies provide us with knowledge? and, if so, (2) Why are the testimonies of experts and local citizens essential in evidence-based policy (EBP)?

The classical definition of knowledge is 'justified true belief' (JTB), which holds that a proposition qualifies as knowledge if, and only if, it is *believed* by someone, is *true*, and is *justified*. Although the inadequacy of JTB was highlighted by Gettier (1963), most

epistemologists continue to regard these three elements as necessary conditions for knowledge (Ichikawa & Steup, 2024). In other words, for something to be considered knowledge, it must at minimum be (1) a belief that is (2) true and (3) justified.

A testimony can certainly satisfy the belief requirement. For instance, if a doctor, after examining me, states that I have Singapore flu, I—already trusting the doctor's authority—can readily accept this diagnosis. This means that the doctor's testimony generates a belief in me regarding my own condition. This belief, based on the doctor's testimony, could also be true if I am indeed infected with enterovirus. Thus, the doctor's testimony has resulted in a true belief. However, is my true belief based on the doctor's testimony (3) justified?

There are two main views on this question. The first view is the reductionist view. This view assumes that testimonial-based beliefs can be justified by other epistemic sources, such as perception, memory, and inference. In other words, beliefs based on testimony cannot be justified by the testimony itself but can only be justified by other epistemic sources that are indeed recognized as reliable. This view can be traced to Hume (O'Brien, 2024) who stated that:

> "[t]he reason, why we place any credit in witnesses and historians, is not derived from any *connexion*, which we perceive *a priori*, between testimony and reality, but because we are accustomed to find a conformity between them," (Hume, 2007, p. 82).

This means that, according to the reductionist view, my belief based on the doctor's testimony is justified not because of the testimony itself, but because I have *perceived* that the person I met is indeed a doctor whom I have known for a long time and I also *remember* that the doctor always tells the truth, so then I *conclude* that what the doctor said when I was examined must also be true. Thus, my testimonial-based belief that I have Singapore flu is justified by perception, memory, and inference all at once.

This reductionist view is usually formulated as follows (Leonard, 2023):

> "A hearer is justified in believing what a speaker says if, and only if, they (a) have positive reasons for thinking that the speaker's testimony is reliable, where these reasons are not themselves ultimately based on testimony, and (b) do not have any undefeated defeaters that indicate that the speaker's testimony is false or unlikely to be true."

So, for a belief based on testimony to be justified, in addition to having justification from an epistemic source other than the testimony itself, the recipient of the testimony must also have no undefeated defeaters that show the testimony is false or unlikely to be true. If, for example, I found an undefeated defeater indicating that the doctor I met was under the influence of alcohol and therefore could have given me a wrong diagnosis, then my belief based on the doctor's testimony is not justified. The presence of such a defeater prevents me from concluding that the doctor is telling the truth, as he has consistently done in the past.

The second view rejects the reductionist view of testimony. According to this non-reductionist view, testimonial-based beliefs can be justified if a single condition is met, namely there is no evidence refuting the accuracy of the testimony (Leonard, 2023). In other words, the recipient of testimony does not need to have other epistemic sources outside of testimony to justify his belief.

The debate between reductionism and non-reductionism regarding testimony remains unresolved. However, this paper does not aim to determine which view is correct or more defensible. Instead, it focuses solely on whether testimonial-based beliefs can be justified. Both reductionists and non-reductionists agree that such beliefs can indeed be justified; they differ only on the source of this justification. Reductionists hold that the justification for testimonial-based beliefs depends on other epistemic sources outside of testimony, whereas non-reductionists argue that no

external sources are necessary—justification is inherent to testimony itself.

Therefore, regardless of which view is ultimately correct, this paper can still defend its thesis that testimony is capable of producing knowledge—at least in the sense of justified true belief (JTB).

Knowledge generated from testimony can serve as supplementary evidence in evidence-based policy formulation. While randomized controlled trials (RCTs) provide empirical evidence for the effectiveness of a policy intervention within a specific study population, implementing this intervention in other populations requires contextual and qualitative knowledge specific to those groups. This additional knowledge helps ensure that a policy intervention effective in one population will also be effective in the target population.

Contextual knowledge offers insight into the social, economic, cultural, and political background of the target population, while qualitative knowledge captures the subjective dimensions of population members, such as their daily experiences, aspirations, motivations, and preferences. Contextual knowledge can be gathered from the testimony of experts, such as anthropologists, sociologists, or historians, who possess specialized understanding of the population. This contextual information, as presented by these experts, should be taken into account prior to implementing a policy intervention. For instance, if this information reveals a misalignment between the proposed policy and the background of the target population, policymakers should consider making adjustments.

Qualitative knowledge about a population can be derived directly from the testimonies of its members, gathered through surveys, aspiration sessions, or public hearings. Such knowledge is crucial in policy formulation because the effectiveness of a policy intervention depends not only on objective factors but also on the motivations, preferences, and other subjective aspects of the target population. Policy formulation rests on the assumption that both the

implementing party and the intended beneficiaries will act in certain ways due to their specific motivations and levels of agency (Le Grand, 2003, p. 2). Therefore, accounting for the subjective dimensions of population members is essential in developing effective public policy.

So, what types of testimony can serve as a foundation for formulating public policies? According to the non-reductionist view, any testimony can be used as a basis for policy, provided there is no undefeated defeater for that testimony. In contrast, the reductionist view requires not only the absence of an undefeated defeater but also additional reasons for relying on the testimony—such as the credibility of the source. For instance, a testimony from a widely respected expert, who is unlikely to risk their reputation by providing false information, would satisfy reductionist criteria. Thus, the non-reductionist criteria for determining the eligibility of testimony as a basis for policy are broader than those of the reductionist view. In this context, we lean toward using reductionist criteria when formulating public policy.

CONCLUSION

In conclusion, while randomized controlled trials (RCTs) provide valuable empirical evidence for policy interventions, they are not sufficient on their own to inform effective public policy. The INUS framework of causation demonstrates that a policy's effectiveness depends on multiple, interrelated factors. Therefore, policymakers require evidence beyond what RCTs alone can provide. Supplementing RCTs with contextual and qualitative knowledge, including the direct testimonies of experts and community members, enables a more comprehensive understanding of the target population. This approach addresses the subjective and contextual nuances that RCTs alone cannot capture, such as motivations, preferences, and cultural factors that influence how a policy may be received and implemented.

Drawing on the reductionist framework, we argue that reliable testimony—especially from credible and respected sources—should

play a foundational role in evidence-based policy. By integrating the reductionist criteria, policymakers ensure that testimonial knowledge used in policy formulation is both relevant and dependable, thereby enhancing the potential for successful policy outcomes. This paper, therefore, supports a balanced approach to evidence-based policy, one that values both empirical evidence and the qualitative, testimonial insights that provide depth and context to policy decisions.

Thus, emphasizing the importance of testimony in public policymaking, we encourage policymakers to adopt the principle "test, learn, and listen" rather than "test, learn, and adapt".

ACKNOWLEDGEMENT

This research was funded by *Program Peningkatan Academic Excellence Skema C* 2024, Universitas Gadjah Mada (6530/UN1.P1/PT.01.03/2024).

REFERENCES

Adams, J., & Hillman, M. (2001). The risk compensation theory and bicycle helmets. *Injury Prevention*, *7*(2), 89–91. https://doi.org/10.1136/ip.7.2.89

Baron, J. (2018). A Brief History of Evidence-Based Policy. *The ANNALS of the American Academy of Political and Social Science*, *678*(1), 40–50. https://doi.org/10.1177/0002716218763128

Birkland, T. A. (2020). *An Introduction to the Policy Process Theories, Concepts, and Models of Public Policy Making* (5th ed.). Routledge.

BPMI Setwapres. (2021, Jan 09). *Apresiasi langkah cepat MUI keluarkan fatwa halal vaksin Sinovac*. Kementerian Sekretariat Negara Republik Indonesia. https://setneg.go.id/baca/index/apresiasi_langkah_cepat_mui_keluarkan_fatwa_halal_vaksin_sinovac

Brooks, G., Burton, M., Cole, P., Miles, J., Torgerson, C., & Torgerson, D. (2008). Randomised controlled trial of incentives to improve attendance at adult literacy classes. *Oxford Review of Education*, *34*(5), 493–504.

https://doi.org/10.1080/03054980701768741

Buck, S., & McGee, J. (2015). *Why Government Needs More Randomized Controlled Trials: Refuting the Myths*.

Canadian Task Force on the Periodic Health Examination. (1979). The Periodic Health Examination: Task Force Report. *Canadian Medical Association Journal*, *121*(9), 1193–1254.

Cartwright, N. (2012). Presidential Address: Will This Policy Work for You? Predicting Effectiveness Better: How Philosophy Helps. *Philosophy of Science*, *79*(5), 973–989. https://doi.org/10.1086/668041

Cartwright, N., & Hardie, J. (2012). *Evidence-Based Policy: A Practical Guide to Doing It Better*. Oxford University Press.

Cartwright, N., & Stegenga, J. (2011). *A Theory of Evidence for Evidence-Based Policy* (P. Dawid, W. Twining, & M. Vasilaki, Eds.). British Academy. https://doi.org/10.5871/bacad/9780197264843.003.0011

Cowen, N., Virk, B., Mascarenhas-Keyes, S., & Cartwright, N. (2017). Randomized Controlled Trials: How Can We Know "What Works"? *Critical Review*, *29*(3), 265–292. https://doi.org/10.1080/08913811.2017.1395223

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, *210*, 2–21. https://doi.org/10.1016/j.socscimed.2017.12.005

Dorsch, M. M., Woodward, A. J., & Somers, R. L. (1987). Do bicycle safety helmets reduce severity of head injury in real crashes? *Accident Analysis & Prevention*, *19*(3), 183–190. https://doi.org/10.1016/0001-4575(87)90002-9

Dye, T. R. (2013). *Understanding Public Policy* (14th ed.). Pearson.

Edwards, P., Arango, M., Balica, L., Cottingham, R., El-Sayed, H., Farrell, B., Fernandes, J., Gogichaisvili, T., Golden, N., Hartzenberg, B., Husain, M., Ulloa, M. I., Jerbi, Z., Khamis, H., Komolafe, E., Laloë, V., Lomas, G., Ludwig, S., Mazairac, G., … Yutthakasemsunt, S. (2005). Final results of MRC CRASH, a randomised placebo-controlled trial of intravenous

corticosteroid in adults with head injury—outcomes at 6 months. *The Lancet*, *365*(9475), 1957–1959. https://doi.org/10.1016/S0140-6736(05)66552-X

Gamoran, A. (2018). Evidence-Based Policy in the Real World: A Cautionary View. *The ANNALS of the American Academy of Political and Social Science*, *678*(1), 180–191. https://doi.org/10.1177/0002716218770138

Gelman, A. (2018). Benefits and limitations of randomized controlled trials: A commentary on Deaton and Cartwright. *Social Science & Medicine*, *210*, 48–49. https://doi.org/10.1016/j.socscimed.2018.04.034

Goldacre, B. (2011). How can you tell if a policy is working? Run a trial. *The Guardian*. https://www.theguardian.com/commentisfree/2011/may/14/bad-science-ben-goldacre-randomised-trials

Haynes, L., Service, O., Goldacre, B., & Torgerson, D. (2012). Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2131581

Hayward, G. C., & Kirst, M. W. (2002). *What have we learned about class size reduction in California?* https://edpolicyinca.org/publications/what-have-we-learned-about-class-size-reduction-california

Howlett, M., & Giest, S. (2015). Policy Cycle. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 288–292). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.75031-8

Hume, D. (2007). *An Enquiry concerning Human Understanding*. Oxford University Press.

Ichikawa, J. J., & Steup, M. (2024). The Analysis of Knowledge. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall). https://plato.stanford.edu/archives/fall2024/entries/knowledge-analysis/

Kemm, J. (2006). The limitations of 'evidence-based' public health. *Journal of Evaluation in Clinical Practice*, *12*(3), 319–324.

https://doi.org/10.1111/j.1365-2753.2006.00600.x

Khosrowi, D., & Reiss, J. (2019). Evidence-Based Policy: The Tension Between the Epistemic and the Normative. *Critical Review*, *31*(2), 179–197. https://doi.org/10.1080/08913811.2019.1688520

Knox, V., Hill, C. J., & Berlin, G. (2018). Can Evidence-Based Policy Ameliorate the Nation's Social Problems? *The ANNALS of the American Academy of Political and Social Science*, *678*(1), 166–179. https://doi.org/10.1177/0002716218769844

La Caze, A., & Colyvan, M. (2017). A Challenge for Evidence-Based Policy. *Axiomathes*, *27*(1), 1–13. https://doi.org/10.1007/s10516-016-9291-5

Le Grand, J. (2003). *Motivation, Agency, and Public Policy*. Oxford University PressOxford. https://doi.org/10.1093/0199266999.001.0001

Leonard, N. (2023). Epistemological Problems of Testimony. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Spring).

Lewis, D. (1973). Causation. *The Journal of Philosophy*, *70*(17), 556–567.

Lisciandra, C. (2020). Reflections on the 2019 Nobel Memorial Prize Awarded to Banerjee, Duflo, and Kremer. *Erasmus Journal for Philosophy and Economics*, *13*(1). https://doi.org/10.23941/ejpe.v13i1.487

Littell, J. H., & White, H. (2018). The Campbell Collaboration. *Research on Social Work Practice*, *28*(1), 6–12. https://doi.org/10.1177/1049731517703748

Mackie, J. L. (1965). Causes and Conditions. *American Philosophical Quarterly*, *2*(4), 245–264.

Menzies, P., & Beebee, H. (2024). Counterfactual Theories of Causation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Summer).

Montuschi, E. (2009). Questions of Evidence in Evidence-Based Policy. *Axiomathes*, *19*(4), 425–439. https://doi.org/10.1007/s10516-009-9085-0

Moore, M. S. (2009). The Counterfactual Theory of Causation. In

Causation and Responsibility: An Essay in Law, Morals, and Metaphysics (pp. 391–425). Oxford University Press.

Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, *5*(2), 113–127.

O'Brien, D. (2024). Hume and the Epistemology of Testimony. In S. Stapleford & V. Wagner (Eds.), *Hume and Contemporary Epistemology* (pp. 315–330). Routledge. https://doi.org/10.4324/9781003394129-26

Parsons, W. (2002). From Muddling Through to Muddling Up - Evidence Based Policy Making and the Modernisation of British Government. *Public Policy and Administration*, *17*(3), 43–60. https://doi.org/10.1177/095207670201700304

Pearce, W., & Raman, S. (2014). The new randomised controlled trials (RCT) movement in public policy: challenges of epistemic governance. *Policy Sciences*, *47*(4), 387–402. https://doi.org/10.1007/s11077-014-9208-3

Pérez-González, S. (2024). Evidence of mechanisms in evidence-based policy. *Studies in History and Philosophy of Science*, *103*, 95–104. https://doi.org/10.1016/j.shpsa.2023.11.006

Peters, B. G. (2010). *American Public Policy: Promise and Performance* (8th ed.). CQ Press.

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., & Lavenberg, J. G. (2013). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency. *Cochrane Database of Systematic Reviews*. https://doi.org/10.1002/14651858.CD002796.pub2

Robinson, D. L. (2006). No clear evidence from countries that have enforced the wearing of helmets. *BMJ*, *332*(7543), 722.2-725. https://doi.org/10.1136/bmj.332.7543.722-a

Sackett, D. L. (1989). Rules of Evidence and Clinical Recommendations on the Use of Antithrombotic Agents. *Chest*, *95*(2), 2S-4S. https://doi.org/10.1378/chest.95.2_Supplement.2S

SIGN. (2011). *SIGN 50: A guideline developer's handbook*.

Smith, A. F. M. (1996). Mad Cows and Ecstasy: Chance and Choice in an Evidence-Based Society. *Journal of the Royal Statistical*

Society. Series A (Statistics in Society), *159*(3), 367. https://doi.org/10.2307/2983324

Thompson, R. S., Rivara, F. P., & Thompson, D. C. (1989). A Case-Control Study of the Effectiveness of Bicycle Safety Helmets. *New England Journal of Medicine*, *320*(21), 1361–1367. https://doi.org/10.1056/NEJM198905253202101

Wasserman, R. C., Waller, J. A., Monty, M. J., Emery, A. B., & Robinson, D. R. (1988). Bicyclists, helmets and head injuries: a rider-based study of helmet use and effectiveness. *American Journal of Public Health*, *78*(9), 1220–1221. https://doi.org/10.2105/AJPH.78.9.1220

Weightman, A., Ellis, S., Cullum, A., Sander, L., & Turley, R. (2005). *Grading evidence and recommendations for public health interventions: developing and piloting a framework*.