Reimagining Evidence-Based Policymaking: A Research Agenda for Making Evidence Useful, Usable, and Used

Elizabeth Linos

Harvard Kennedy School, Harvard University, Cambridge, MA 02138

Abstract

Despite growing investments in evidence-based policymaking, a significant gap remains between producing policy-relevant research and achieving policy-impactful results. This article argues that overcoming this gap requires reimagining evidence-based policymaking as a series of informational, translational, and organizational hurdles that policymakers face, deserving the same rigorous academic attention as the production of evidence itself. By coalescing around an interdisciplinary research agenda that aims to better understand – and reduce – these barriers, we can make evidence more useful, usable, and used. In doing so, we can move evidence-based policymaking beyond theoretical relevance to deliver meaningful social impact.

The past two decades have seen a surge in efforts to integrate data and evidence into government. Policymakers and academics are increasingly focused on producing actionable evidence in real-world settings, with randomized controlled trials (RCTs) playing a key role in testing what works – and what does not – across a wide range of contexts (Abdul-Razzak & Hallberg, 2024; Banerjee et al., 2018; Bhatt et al., 2023; Dube et al., 2023; Heller et al., 2017; Kappes et al., 2023; Milkman et al., 2022; Toma et al., 2023). In the United States, these efforts have been ensconced in law in the form of the 2018 Evidence Act, which requires federal agencies to prioritize evidence-building. Globally, initiatives like the United Kingdom's What Works Network and the OECD's evidence-informed policy guidelines reflect similar commitments to embedding evidence in decision-making systems (Bristow et al., 2015; ESRC, 2016; Gough et al, 2018; OECD, 2024).

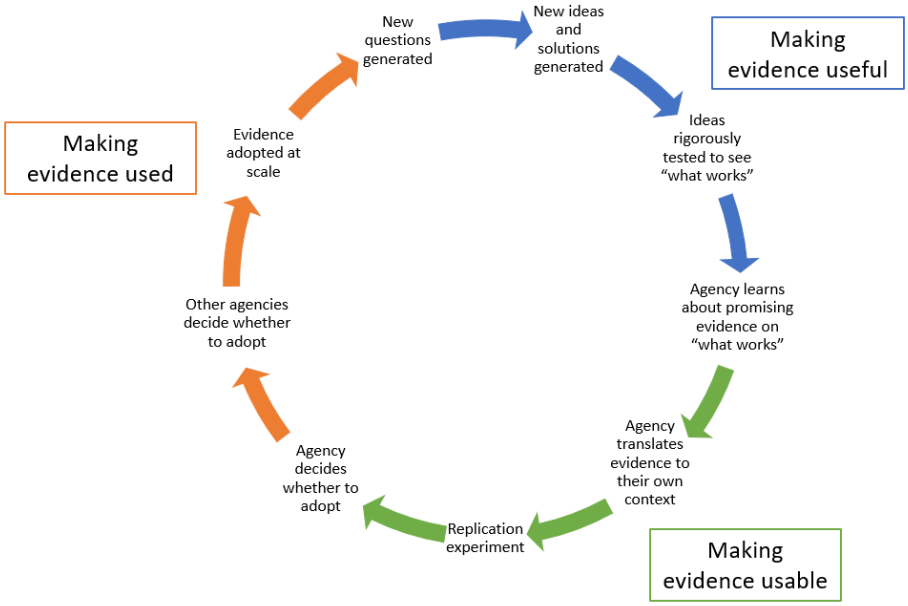
And yet, most government spending is not guided by evidence-based programs or policies (Bridgeland & Orszag, 2013). A growing anti-science rhetoric in politics and a mistrust of academic insights in policymaking exacerbate this gap (Pearson, 2024). And an interdisciplinary literature documents how difficult it is to take academic evidence to scale in policy settings (e.g., Bogenschneider & Corbett, 2010; Haines & Donald, 1998; Kajermo et al. 2010). For example, Kremer et al. (2021) note that the evidence behind USAID-funded RCTs was only adopted at scale in less than a third of cases, with Rao (2024) finding no association between a country's program evaluations and policy spending. The same pattern holds true in the US: DellaVigna et al. (2024) document an adoption rate of approximately 30% in behavioral interventions conducted by US cities, and find that the strength of the evidence is not a predictor of adoption.

The challenge is clear: while the academic community is producing policy-relevant research—research with clear policy implications—we are falling short of producing policy-impactful research that changes which programs are adopted and how services are delivered at scale. Understanding why this gap persists and how to close it is essential for ensuring that the evidence-based policymaking movement meaningfully improves societal outcomes. This is, in itself, a pressing policy-relevant and policy-impact research question.

To realize its potential, evidence-based policymaking must be re-imagined as a series of behavioral hurdles, each demanding the same level of research attention and rigor that produced the original evidence base (see Figure 1). Taking the perspective of the user – in this case the public servant – we need to better understand the underlying barriers at each step in this process: What makes it hard for a public servant to know what promising evidence exists, whether it is relevant in their context, how to translate findings into a given program or service, and how to adopt best practice at scale within a bureaucracy? These questions do not lend themselves to a disciplinary focus. Indeed, there exist studies in political science, organizational economics, behavioral science, implementation science, public management, and sociology that examine different facets of this problem. But much of the existing literature focuses on describing the problem. For instance, a host of new studies examine the predictors of evidence use and

adoption, as well as how policymakers (and politicians) form and update their beliefs about evidence. Building on this important foundation, we must now move beyond descriptives; we need a comprehensive and interdisciplinary research agenda aimed at identifying and reducing these hurdles to make evidence more useful, usable, and used.

Figure 1. The evidence-based policymaking cycle



Making Evidence Useful

The first series of hurdles in evidence-based policymaking is informational: policymakers need to know that evidence exists, understand its findings and its level of rigor to be appropriately confident in the results, and ultimately value it as useful. Policymakers are, of course, humans like everybody else, and all the behavioral biases that may affect how an individual consumes new information – bounded rationality, inattention, ambiguity aversion, status quo bias, availability bias and more – will also apply to policymakers and practitioners (e.g., Battaglio et al., 2018; Bellé et al., 2017; and Moynihan & Lavertu, 2012 for examples from policy contexts). As such, all the tools that behavioral scientists have developed to improve attention and understanding, such as simplifying and shortening messages, adjusting the messenger, or personalizing communications, should apply here. Recent work suggests that this is the case: the format and presentation of evidence significantly affect how policymakers engage with findings. For example, policymakers are more likely to engage with evidence when it is concise,

accessible, and tailored to their context (Hjort et al., 2021; Nakajima, 2021). Similarly, policymakers are more likely to update their beliefs when the messenger is credible and the topic is politically salient (Henig, 2009; Peterson, 2018).

Yet, we do not know whether knowing and understanding evidence makes it more useful to policymakers, even if it leads them to update their priors on "what works." This theory of change presumes that policymakers will value evidence more if they understand it better. Whether this is, in fact, the case is an empirical question. The extant literature documents wide variation among policymakers in whether research-based evidence is considered valuable to begin with. On the one hand, Toma and Bell (2024) note that, at baseline, policymakers' preferences are relatively inelastic to evidence, meaning their assessment of the value of a program does not change substantially based on program impact measures. Hjort et al. (2021), on the other hand, find that mayors are willing to pay to learn the findings of an evaluation, suggesting some demand for evidence, although their sample focuses on elected officials. Bogenschneider and Corbett (2010) similarly interviewed almost 200 US State legislators and found that legislators range from "enthusiastic *users* of research" to "enthusiastic *nonusers* of research," with a full third of legislators falling into the latter camp. An open question is if this ratio has shifted further away from science-oriented policymakers in the current political climate, and whether and how this varies for elected officials versus public managers. But regardless of the ratio, taking evidence to scale means better understanding and overcoming hurdles even for those policymakers who would not a-priori see the use of evidence as useful.

The vast majority of research on evidence-based policymaking stops at thinking through *how* to present existing research more effectively. A crucial next step is to ask *what* to present to make evidence more useful. Imagine a world where all research is presented in terms of effect sizes and statistical significance – even with clear, concise, and behaviorally-informed text and images – yet policymakers' judgment on whether this information is useful depends on different components of the research that may speak to other political, budgetary, or capacity constraints they face: How much did it cost? What are the heterogeneous effects for specific subpopulations that are politically salient? How related is the outcome measured in the study to the outcome residents are monitoring? What is the likelihood that communities will support this policy? How many times has this new approach been tested and where? Emerging evidence suggests that policymakers regularly trade-off impact measures with other factors they value such as local proximity or constituent preferences (Carattini et al, 2024; Vivalt et al., 2024), but summaries of results from academic research, even in policy-oriented clearinghouses, rarely include the implementation or political economy details that policymakers may need to consider when making decisions. As such, we do not know whether presenting this additional information would shift perceptions of usefulness, nor how policymakers weigh all of the different factors that could influence the use of evidence.

This reality points to an urgent need for more research on three key questions: (a) Does varying *how* information is presented change demand for its use, particularly in different political contexts and for heterogeneous populations of policymakers and practitioners? (b) How does varying *what* information is presented change what evidence is perceived as more useful, in what contexts, and for whom? And (c) in an environment where policymakers face constraints, what are the relative weights that policy makers place on various components of evidence when deciding what evidence is most valuable to consider? More broadly, is evidence more useful when it answers questions that a government has actually asked?

Making Evidence Usable

The second series of challenges is translational: even when evidence is deemed useful, policymakers may face barriers in translating the insights from one context into their own. Translating findings is difficult even in purely academic circles. Camerer et al. (2018) for example, show that only 13 out of 21 social science experiments in Science or Nature replicate, with the effect sizes being half as large as the original studies, in line with research on the "voltage drop" (Kilbourne et al., 2007; List, 2022). Literature on the "transportability" problem also notes the various conditions under which insights from an experiment in one setting may (not) apply in other contexts (Bareinboim & Pearl, 2013). When we ask policymakers to "translate evidence into policy" – to make it usable in their own context – we are asking for something much more difficult. Rather than simply repeating the same intervention in a similar-enough context (with an emphasis on fidelity to the original model), we ask policymakers to essentially produce a "conceptual replication" – parsing out the mechanism behind an intervention, making an educated guess on whether the same mechanism would apply to a new population and outcome, and using those hypotheses to adapt existing evidence into their own context and take the idea to scale. This requires understanding not simply "what works", but also *why* and *how* it works.

Two specific challenges emerge. First, successful translation of evidence requires understanding the conditions in which it was generated. Implementation science has underscored the critical importance of context, design decisions, organizational structures, and other systemic factors in shaping the outcomes of interventions. Studies have documented how even small changes in these variables can significantly alter effectiveness (e.g., Damschroder et al., 2009; Proctor et al., 2011). Even in cases where an intervention has been rigorously tested with the gold standard of evaluation, a randomized controlled trial, only a small number of interventions have been replicated multiple times in different contexts, and an even smaller subset have been replicated outside of WEIRD (Western, Educated, Industrialized, Rich, and Democratic) settings (Rad et al., 2018). This matters both because a literature on "site selection bias" points to a positive correlation between sites used early in experimentation and effect sizes (Allcott, 2015; Wang & Yang, 2021), and because ideas that have only been tested in a very narrow set of policy contexts

may limit a policymaker's ability to ascertain whether a given intervention may be usable in their own context.

Second, recent debates on the average effects of behavioral interventions have shed more light on the continued challenge of publication bias in academia (e.g., Andrews & Kasy, 2019; Brodeur et al., 2016; Simonsohn et al., 2014). If the only evidence that is published or shared widely is surprising success stories (that may or may not replicate in other contexts), policymakers may expect larger effects when they translate these findings into their own contexts (DellaVigna & Linos, 2022). This could create a long-term barrier to evidence being usable: When interventions are implemented and the results prove to have a lower-than-expected effect, this could create negative feedback loops for future trust in – and use of – both experimentation and evidence (Briscese & List, 2024).

The sobering implication of these challenges is that policymakers cannot – and should not – expect similar effects when they try to translate evidence-based interventions into their own contexts, especially if some part of the delivery of the intervention or the target population is fundamentally different than the original study. Instead, until we understand *what works, for whom,* and *when* across a host of policy settings, best practice would suggest that policymakers rigorously pilot test any new intervention in their own context.

This requires a research agenda that seeks to first understand how to build demand for and reduce barriers to experimentation – not just evidence adoption – both amongst academics and policymakers. In an ideal world, this would also involve more coordination across sites for field experiments on the same issue, in an approach used by the MetaKeta Initiative and others (Dunning et al., 2019). We know surprisingly little about what levels of support and internal capacity or what types of interventions are more likely to create demand for experimentation and replication in government, especially across sociopolitical contexts, and even less about how to shift these capacity constraints.

Second, we need to ask what factors influence replicability of tested interventions across contexts. Traditional fidelity-focused models of replication emphasize adherence to a rigid protocol in order to understand effects at scale. A research agenda focused on improving evidence-based policymaking may need to instead consider an approach that prioritizes adaptation within evolving contexts (Aarons et al., 2012; Chambers et al., 2013). These models allow for greater flexibility and responsiveness, but raise important empirical questions such as: Does allowing for more flexibility in implementation conditions lead to higher translation of evidence? Does the translated version maintain similar effect sizes as the original study? And, ultimately, does this approach make existing evidence more usable?

Making Evidence Used

The third series of challenges is organizational: even when political actors know and value evidence as useful, and have translated some evidence into their context through a pilot experiment, there are a number of barriers to sustained adoption at scale. Dellavigna et al. (2024) document that about 30% of "usable" interventions are in fact *used* by the same cities that tested them. In other words, less than one-third of city departments in the US that experimentally tested evidence-based interventions ultimately ended up adopting them in the long-run. The same patterns likely hold true in private and non-profit sectors as well (e.g., Athey & Luca, 2019; List, 2022).

A broad literature points to a number of barriers that may influence organizations' ability to adopt evidence, even when it is useful. First, some organizations may not have the routines, practices, or managerial structures that would make them "learning organizations" (Moynihan & Landuyt, 2009) or may not have the informal relational contracts and manager-employee trust that underpin adoption of many best practices (Gibbons & Henderson, 2012). Other literature examines how a lack of "organizational slack" - determined by size, wealth, and personnel – may limit innovation and the use of evidence (e.g., Argote & Miron-Spektor, 2011; Besley & Persson, 2009; de Vries et al., 2015). In particular, larger or wealthier organizations may, almost by definition, have more resources to be able to adopt new evidence (Fernandez & Wise, 2010; Naranjo-Gil, 2009). But little is known about what strategies or approaches can effectively shift organizational practices to increase evidence use.

A second barrier to adoption is organizational inertia or a tendency to maintain the status quo. DellaVigna et al. (2024) found that the main predictor of adoption was whether the treatment was tweaking an existing process. Put differently, in the sample of cities studied, incremental improvements to existing infrastructure were much more likely to be adopted than interventions that required new processes. This may suggest a need to shift how we think about the trade-off between different types of interventions. For instance, rather than prioritizing interventions that will give us the largest impact on policy outcomes in the abstract, we may want to measure policy success by likelihood (or feasibility) of adoption at scale, when designing interventions for practice, that may make lighter-touch tweaks to the status quo more appealing. But the underlying mechanisms are not yet well understood. It could be the case that incremental improvements are more likely to be adopted simply because of organizational capacity: it typically requires more resources to adopt a new process than modify an existing one. But it is also possible that adopting a new process requires stakeholder buy-in that is not required for incremental "tweaks." This points to a third potential barrier to adoption: political context.

The political environment of an organization can influence its willingness and ability to use evidence. Policymakers do not make decisions in a vacuum and a single public servant or political leader rarely has the ability within an organization to "adopt" evidence. A large

literature on policy diffusion points to many factors that influence political buy-in for an idea, including relationships between political leaders, windows of opportunity, or key policy champions (Baumgartner & Jones, 1993; Boushey, 2012; Oliver et al., 2014; Shipan & Volden, 2008). But there is little evidence on how to build buy-in within a political context or agency, nor on whose buy-in is most consequential when it comes to evidence use or how this differs by organizational factors.

More broadly, although we have a growing body of compelling research on what predicts evidence use and adoption, we have a dearth of experimental research on how to influence this process. This would require moving beyond studies that test how to reduce information, knowledge, and experimentation barriers, in order to evaluate – with the same rigor that is used to test the original intervention -- what works in supporting organizations to *adopt* evidence. The time for these evaluations is now. Building on previous investments that fostered the evidence base around "what works" for critical policy outcomes, large investments are now being made in strategies that are commonly believed to motivate evidence use: hundreds of millions of dollars are spent on trainings, technical assistance, and clearinghouses, all aimed at increasing adoption of best practice or evidence. But these approaches are rarely tested empirically, nor are they tested against each other.

This underscores the pressing need for research on what works to increase actual adoption of evidence including: (a) What types of support for policymakers increases adoption? (b) How can existing networks be leveraged to spread evidence and increase adoption among members? And (c) what types of infrastructure and human capital investments in government shift the ability of policymakers to adopt evidence at scale?

Conclusion

We are still in the early stages of understanding how evidence can inform practice (as well as how practice can inform evidence) and, in particular, how to effectively bridge the research-policy divide. Making government more evidence-based is not simply a matter of giving the right people the right information. In the same way that researchers have known for decades that telling people they should go to the gym doesn't magically lead to consistent exercise habits, the gap between presenting the evidence and ensuring its sustained use at scale is inherently a behavioral research question. The challenge is also not a one-stage problem; it involves overcoming a series of hurdles that will require academics, policymakers, and the larger evidence-based policymaking ecosystem to shift their behavior.

In academia, we need to start seeing the use of our own research findings – particularly its ability to be translated into new contexts and understood by a wide range of actors – as an important research question in itself, rather than afterthought or something left to the purview of communication teams. This may require setting up new models for research-practitioner

collaborations that allow academics to be directly responsive to urgent policy priorities while still conducting rigorous research that contributes to generalizable knowledge. It may also require researchers to collect and present other types of information in publications, as well as a broader incentive structure that both allows for and encourages more replications and coordinated experimentation in cohorts.

In policy and practitioner communities, while there is a growing investment in conducting rigorous experimentation in government, the results we see are limited to a very small subset of government agencies. Importantly, even amongst those early innovators who have conducted experiments, the next stage of this process is for policymakers to invest in *adoption* of evidence up-front. This could include building commitment tools or other mechanisms through which a new piece of evidence is adopted, should it prove to be effective in testing. It also requires investments in the people, processes, and platforms in government that will make adoption more seamless, once evidence exists. Philanthropic organizations and other funders will play a critical role in determining whether these questions of implementation and adoption are considered equally valuable behavioral questions as more traditional behavioral economics questions, and they have a particular role to play in investing in innovation capacity within government as opposed to building workarounds.

Last, the broader ecosystem of evidence-based policy practitioners – that includes researchers, practitioners, as well as "bridge" organizations, consultants, advocacy groups, and funders – are critical to ensuring that evidence is truly relevant to policy. As a community, we must commit to transparency of results, regardless of the outcome, and encourage comfort with small (but real) effects. Crucially, actively expanding who is involved in the production of evidence, both by including more voices in the co-design of interventions and in the selection of outcome metrics, may be critical in creating more trust in science and more trustworthy science. Now, more than ever, as growing anti-science rhetoric threatens to further separate data and evidence from policymaking, we have an obligation to reimagine the process of understanding, translating, and adopting evidence as a series of behavioral pain points that warrant the same level of academic study and inquiry as the evidence production to begin with. If we pursue this interdisciplinary research agenda, we will ensure that the evidence-based policymaking movement meets its potential: better policy outcomes through facts.

References

Aarons, G. A., Green, A. E., Palinkas, L. A., Self-Brown, S., Whitaker, D. J., Lutzker, J. R., Silovsky, J. F., Hecht, D. B., & Chaffin, M. J. (2012). Dynamic adaptation process to implement an evidence-based child maltreatment intervention. *Implementation Science, 7*, 32. https://doi.org/10.1186/1748-5908-7-32

Abdul-Razzak, N., & Hallberg, K. (2024). *Unpacking the impacts of a youth behavioral health intervention: Experimental evidence from Chicago*. Annenberg Institute at Brown University. https://doi.org/10.26300/b9dh-vq42

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics, 130*(3), 1117–1165. https://doi.org/10.1093/qje/qjv015

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review, 109*(8), 2766–94. https://doi.org/10.1257/aer.20180310

Argote, L., & Miron-Spektor, E. (2011). Organizational learning: From experience to knowledge. *Organization Science, 22*(5), 1123–1137. https://doi.org/10.1287/orsc.1100.0621

Athey, S., & Luca, M. (2019). Economists (and Economics) in Tech Companies. *Journal of Economic Perspectives, 33*, 209–230. https://doi.org/10.1257/jep.33.1.209.

Banerjee, A., Hanna, R., Kyle, J., Olken, B. A., & Sumarto, S. (2018). Tangible information and citizen empowerment: Identification cards and food subsidy programs in Indonesia. *Journal of Political Economy, 126*(2), 451–491. https://doi.org/10.1086/696226

Bareinboim, E., & Pearl, J. (2013). A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference, 1*(1), 107–134. https://doi.org/10.1515/jci-2012-0004

Battaglio, R. P., Belardinelli, P., Bellé, N., & Cantarelli, P. (2018). Behavioral Public Administration *ad fontes*: a synthesis of research on bounded rationality, cognitive biases, and nudging in public organizations. *Public Administration Review, 79*(3), 304–320. https://doi.org/10.1111/puar.12994

Baumgartner, F. R., & Jones, B. D. (1993). *Agendas and Instability in American Politics.* University of Chicago Press.

Bellé, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive biases in performance appraisal: Experimental evidence on anchoring and halo effects with public sector managers and employees. *Review of Public Personnel Administration, 37*(3), 275–294. https://doi.org/10.1177/0734371x17704891

Besley, T., & Persson, T. (2009). The origins of state capacity: Property rights, taxation, and politics. *American Economic Review, 99*(4), 1218–1244. https://doi.org/10.1257/aer.99.4.1218

Bogenschneider, K., & Corbett, T. I. (2010). *Evidence-Based Policymaking: Insights from Policy-Minded Researchers and Research-Minded Policymakers*. Routledge.

Bhatt, M., Heller, S., Kapustin, M., Bertrand, M., & Blattman, C. (2023). *Predicting and Preventing Gun Violence: An Experimental Evaluation of READI Chicago*. https://doi.org/10.31235/osf.io/dks29

Boushey, G. (2012). Punctuated equilibrium theory and the diffusion of innovations. *Policy Studies Journal, 40*(1), 127–146. https://doi.org/10.1111/j.1541-0072.2011.00437.x

Bridgeland, J., & Orszag, P. (2013). *Can government play moneyball?* The Atlantic. https://www.theatlantic.com/magazine/archive/2013/07/can-government-play-moneyball/309389

Briscese, G., & List, J. (2024). Toward an understanding of the political economy of using field experiments in policymaking. *Natural Field Experiments, 799*. The Field Experiments Website. https://ideas.repec.org/p/feb/natura/00799.html

Bristow, D., Carter, L., & Martin, S. (2015). Using evidence to improve policy and practice: the UK What Works Centres. *Contemporary Social Science, 10*(2), 126–137. https://doi.org/10.1080/21582041.2015.1061688

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The Empirics Strike Back. *American Economic Journal: Applied Economics, 8*(1), 1–32. https://doi.org/10.1257/app.20150044

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Carattini, S., Dur, R., & List, J. (2024). Policy evaluation and the causal analysis of public support. *Science, 386*(6721), 490-492. https://doi.org/10.1126/science.adp7497

Chambers, D.A., Glasgow, R.E. & Stange, K.C. (2013). The dynamic sustainability framework: Addressing the paradox of sustainment amid ongoing change. *Implementation Science 8*, 117. https://doi.org/10.1186/1748-5908-8-117

Damschroder, L. J., Aron, D. C., Keith, R. E., Kirsh, S. R., Alexander, J. A., & Lowery, J. C. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50. https://doi.org/10.1186/1748-5908-4-50

DellaVigna, S., Kim, W., & Linos, E. (2024). Bottlenecks for evidence adoption. *Journal of Political Economy, 132*(8), 2748–2769. https://doi.org/10.1086/729447

DellaVigna, S., & Linos, E. (2022). RCTs to scale: Comprehensive evidence from two nudge units. *Econometrica, 90*(1), 81–116. https://doi.org/10.3982/ecta18709

De Vries, H. A., Bekkers, V.J.J.M., & Tummers, L. G. (2015). Innovation in the public sector: A systematic review and future research agenda. *Public Administration, 94*(1), 146–166. https://doi.org/10.1111/padm.12209

Dube, O., MacArthur, S. J., & Shah, A. K. (2023). *A cognitive view of policing*. National Bureau of Economic Research. https://doi.org/10.3386/w31651. Working Paper.

Dunning, T., Grossman, G., Humphreys, M., Hyde, S. D., McIntosh, C., Nellis, G., Adida, C. L., Arias, E., Bicalho, C., Boas, T. C., Buntaine, M. T., Chauchard, S., Chowdhury, A., Gottlieb, J., Hidalgo, F. D., Holmlund, M., Jablonski, R., Kramon, E., Larreguy, H., ... Sircar, N. (2019). Voter information campaigns and political accountability: Cumulative findings from a preregistered meta-analysis of coordinated trials. *Science Advances, 5*(7), eaaw2612. https://doi.org/10.1126/sciadv.aaw2612

ESRC. (2016). *ESRC What Works strategic review: Report of stakeholder survey and documentary analysis.* Economic and Social Research Council. https://www.ukri.org/wp-content/uploads/2017/02/ESRC-220311-WhatWorks-StrategicReviewReport-170202.pdf

Fernández, S., & Wise, L. R. (2010). An exploration of why public organizations 'ingest' innovations. *Public Administration, 88*(4), 979–998. https://doi.org/10.1111/j.1467-9299.2010.01857.x

Gibbons, R., & Henderson, R. (2012). Relational contracts and organizational capabilities. *Organization Science, 23*(5), 1350–1364. https://doi.org/10.1287/orsc.1110.0715

Gough, D., Maidment, C., & Sharples, J. (2018). *UK What Works Centres: Aims, methods and contexts*. EPPI Centre, Social Science Research Unit, UCL Institute of Education, University College London. https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3731

Haines, A., & Donald, A. (1998). Making better use of research findings. *British Medical Journal (BMJ), 317*(7150), 72–75. https://doi.org/10.1136/bmj.317.7150.72

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., & Pollack, H. A. (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics, 132*(1), 1–54. https://doi.org/10.1093/qje/qjw033

Henig, J. R. (2009). *Spin Cycle: How Research Gets Used in Policy Debates--The Case of Charter Schools*. Russell Sage Foundation.

Hjort, J., Moreira, D., Rao, G., & Santini, J. F. (2021). How research affects policy: Experimental evidence from 2,150 Brazilian municipalities. *American Economic Review, 111*(5), 1442–1480. https://doi.org/10.1257/aer.20190830

Kajermo, K. N., Boström, A. M., Thompson, D. S., Hutchinson, A. M., Estabrooks, C. A., & Wallin, L. (2010). The BARRIERS scale -- the barriers to research utilization scale: A systematic review. *Implementation Science, 5*(1), 1–22. https://doi.org/10.1186/1748-5908-5-32

Kappes, H. B., Toma, M., Balu, R., Burnett, R., Chen, N., Johnson, R., Leight, J., Omer, S. B., Safran, E., Steffel, M., Trump, K.-S., Yokum, D., & Debroy, P. (2023). Using communication to boost vaccination: Lessons for COVID-19 from evaluations of eight large-scale programs to promote routine vaccinations. *Behavioral Science & Policy*, *9*(1), 11–24. https://doi.org/10.1177/23794607231192690

Kilbourne, A. M., Neumann, M. S., Pincus H. A., Bauer M. S., & Stall R. (2007). Implementing evidence-based interventions in health care: Application of the replicating effective programs framework. *Implementation Science, 2*, 42. https://doi.org/10.1186/1748-5908-2-42

Kremer, M., Thomas, M., Gallant, S., & Rostapshova, O. (2021). *Is development innovation a good investment? Evidence on scaling and social returns from USAID's innovation fund.* Working Paper.

List, J. A. (2022). *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale.* Crown Currency.

Milkman, K. L., Gandhi, L., Patel, M. S., Graci, H. N., Gromet, D. M., Ho, H., Kay, J. S., Lee, T. W., Rothschild, J., Bogard, J. E., Brody, I., Chabris, C. F., Chang, E., Chapman, G. B., Dannals, J. E., Goldstein, N. J., Goren, A., Hershfield, H., Hirsch, A., Hmurovic, J., ... Duckworth, A. L. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences, 119*(6), e2115126119. https://doi.org/10.1073/pnas.2115126119

Moynihan, D. P., & Lavertu, S. (2012). Does involvement in performance management routines encourage performance information use? Evaluating GPRA and PART. *Public Administration Review, 72*(4), 592–602. https://doi.org/10.1111/j.1540-6210.2011.02539.x

Moynihan, D. P., & Landuyt, N. (2009). How do public organizations learn? Bridging cultural and structural perspectives. *Public Administration Review, 69*(6), 1097–1105. https://doi.org/10.1111/j.1540-6210.2009.02067.x

Nakajima, N. (2021). *Evidence-Based Decisions and Education Policymakers*. https://nozominakajima.github.io/files/nakajima_policymaker.pdf. Working Paper.

Naranjo-Gil, D. (2009). Management information systems and strategic performances: The role of top team composition. *International Journal of Information Management, 29*(2), 104–110. https://doi.org/10.1016/j.ijinfomgt.2008.05.009

OECD. (2024). *LOGIC: Good practice principles for mainstreaming behavioural public policy*. OECD Publishing. https://doi.org/10.1787/6cb52de2-en

Oliver, K., Innvar, S., Lorenc, T., Woodman, J., & Thomas, J. (2014). A systematic review of barriers to and facilitators of the use of evidence by policymakers. *BMC Health Services Research, 14*, 2. https://doi.org/10.1186/1472-6963-14-2

Pearson, H. (2024). Science could solve some of the world's biggest problems. Why aren't governments using it? *Nature, 636*, 26–30. https://doi.org/10.1038/d41586-024-03906-0

Peterson, M. A. (2018). In the shadow of politics: The pathways of research evidence to health policy making. *Journal of Health Politics, Policy and Law, 43*(3), 341–376. https://doi.org/10.1215/03616878-4366136

Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research, 38*(2), 65–76. https://doi.org/10.1007/s10488-010-0319-7

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of *Homo sapiens*: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Rao, M. (2024). *Program evaluations and policy spending: Examining conditional cash transfers in Latin America and the Caribbean.* London School of Economics and Political Science. Working Paper.

Shipan, C. R., & Volden, C. (2008). The mechanisms of policy diffusion. *American Journal of Political Science, 52*(4), 840–857. https://doi.org/10.1111/j.1540-5907.2008.00346.x

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*(2), 534. https://doi.org/10.1037/a0033242

Toma, M., & Bell, E. (2024). Understanding and increasing policymakers' sensitivity to program impact. *Journal of Public Economics, 234*, 105096. https://doi.org/10.1016/j.jpubeco.2024.105096

Toma, M., Burnett, R., Debroy, P., Dimant, E., Liu, J., Safran, E., Say, U., & Schultz, B. (2023). Promoting safe opioid disposal: Experimental evidence on behavioral messaging with financial incentives. *SSRN*. https://doi.org/10.2139/ssrn.4502634

Vivalt, E., Coville, A., & KC, S. (2024). Local knowledge, formal evidence, and policy decisions. *Journal of Development Economics, 174*, 103425. https://doi.org/10.1016/j.jdeveco.2024.103425

Wang, S., & Yang, D. Y. (2021). *Policy experimentation in China: The political economy of policy learning*. National Bureau of Economic Research. https://doi.org/10.3386/w29402. Working Paper.