Local Knowledge, Formal Evidence, and Policy Decisions

Eva Vivalt*

Aidan Coville†

Sampada KC‡

February 3, 2024

Abstract

How do policymakers value advice from local experts versus formal evidence from impact evaluations when making policy decisions? Using a discrete choice experiment run with the World Bank and Inter-American Development Bank, we show that policymakers were willing to accept a program that had a 5.4 percentage point smaller estimated effect on enrollment rates if it were recommended by a local expert - larger than the effects of most programs. We find a similar premium (6.4 percentage points) for impact evaluation evidence as long as it is available from the same country, highlighting the importance policymakers place on local evidence.

*Department of Economics, University of Toronto, eva.vivalt@utoronto.ca

†Development Impact (DIME), World Bank, acoville@worldbank.org

‡University of British Columbia, samkc.student@ubc.ca. We thank Oscar Mitnik, Sebastian Martinez, and Silvia Velez Caroco for enabling the surveys to be run. Antonio Ortiz-Mena and Young Ju Nam provided helpful research assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

1 Introduction

The significant increase in impact evaluations in recent years means that policymakers have more formal evidence than ever before to inform decisions on which policies to pursue. At the same time, policymakers often seem to rely on the recommendations of local experts and there is significant disagreement regarding how well local experts tend to forecast the effects of different programs (e.g., Bessone et al., 2021; Milkman et al., 2022; Iacovone et al., 2023). We consider how much policymakers weigh the recommendations of local experts relative to how much they weigh impact evaluation results and whether there are any features of impact evaluations that make them more likely to influence policymaker decisions, using a discrete choice experiment. We find that policymakers place significant weight on advice from local experts, but they also value evidence from impact evaluations in their country. This highlights the importance of generating evidence that is perceived as contextually relevant if impact evaluations are to inform policy decisions.

We conduct the discrete choice experiment with policymakers and policy practitioners at World Bank (WB) and Inter-American Development Bank (IDB) impact evaluation workshops. The WB workshops are meant to serve as "matchmaking" events that connect government officials interested in impact evaluation with researchers supporting the design of an impact evaluation for one of their programs. The IDB workshops are intended to provide training to improve policymaker and policy practitioners' awareness and understanding of impact evaluations. Participants at both types of workshops are demonstrably interested in impact evaluation, while also being at the frontline of program and policy decisions, making them a particularly relevant target group for this study. Participants include monitoring and evaluation specialists and program officers in charge of a particular program from a low or middle-income country government (*policymakers*) and staff and technical advisors at international organizations or aid agencies (*policy practitioners*). We refer to these two groups jointly as "policy professionals".

We surveyed 190 eligible attendees, providing them with descriptions of programs that are each associated with an impact evaluation result and asking them to choose between them. The impact evaluations attached to the programs differ by their identification strategy; location; impact; and the precision of the estimate. We compare how participants weigh these impact evaluation results relative to advice from a local expert.

We find that policymakers place a relatively high weight on contextual factors, such as whether a local expert recommended the program and whether the program was evaluated by an impact evaluation in their country. This may be due to concerns with how much rigorous evidence from one setting may be transportable to another (Pritchett and Sandefur, 2015; Vivalt, 2020), an important consideration when papers are still fairly geographically clustered (Leight, 2022).

Our study builds on existing work exploring the decision-making processes of policymakers. Rogger and Somani (2023) find that bureaucrats in Ethiopia have beliefs about their constituencies that vary significantly from official statistics and that providing evidence briefings can help reduce this gap. Nellis et al. (2019) consider how policy practitioners weigh meta-analysis results compared to results from individual studies. Toma and Bell (2022) examine the impact of decision aids on policy choice. Hjort et al. (2019) consider Brazilian mayors' willingness-to-pay for information from impact evaluations. Finally, in a companion paper, we consider which attributes policymakers, policy practitioners and researchers find attractive when searching for evidence (Vivalt et al., 2023).

In contrast to this past literature, we ask policymakers to select *programs* rather than *studies*. Seeking out new evidence and weighing evidence when selecting a program are related but conceptually distinct. For example, policymakers could seek information from an impact evaluation that found that a program had a negligible effect in order to investigate the reasons for the lack of impact. However, all else equal, they would not want to choose a program with negligible effects when weighing which program to select. While our experiment considers hypothetical choices, our focus on programs rather than studies is perhaps closer to the main item of interest, namely, how policy professionals make policy decisions.

Further, our short paper is the first to consider how policymakers weigh expert advice relative to impact evaluation results. This is very relevant to the practical concerns of many working in development, as expert knowledge is a resource commonly used by policymakers (Morgan, 2014). More broadly, this paper relates to the concept of "tacit" vs. "explicit" knowledge, a distinction often made in the field of knowledge management. Tacit knowledge can be defined as knowledge held by an individual that is hard to formalize. It may be based on personal experiences and intuitions (Polanyi, 2009). In our experiment, impact evaluation results represent explicit knowledge - they are formalized and can be communicated and understood by others. However, advice from a local expert could include elements of both tacit and explicit knowledge. For example, a local expert could be drawing on a number of formal sources of evidence in coming to their conclusions, and how they interpret past findings and integrate this information with their broader understanding and experience to form a recommendation involves their tacit knowledge. While tacit knowledge has been shown to be important in decision-making in a number of diverse fields (e.g., Podgórski, 2010; Hanna et al., 2014; Meisch et al., 2022), it is relatively understudied in economics, and this is the first paper to examine how policymakers weigh this type of evidence.

Finally, to quantify these trade offs, we consider how much participants would be willing to give up in terms of estimated impact in exchange for a program being supported by a certain kind of information. Program impacts provide a natural unit of analysis for assessing trade-offs because they are analogous to a public budget: they are a real cost born by the public that depends on the choices of the policymaker. Our experimental design allows us to say, for example, that policymakers would accept a program that was not recommended by a local expert, over one that was, only if such a program had at least a 5.4 percentage point higher estimated impact on enrollment rates; further, they would prefer a program evaluated in a different region over one evaluated in their country only if the program evaluated in a different region had at least a 6.4 percentage point higher estimated impact. These estimated impacts are very large compared to the typical effects of popular programs that improve enrollment rates.[1]

Our results highlight the importance of local knowledge for policy decisions and suggest that researchers looking to maximize their impact leverage appropriate settings and communicate their research findings to those local experts from whom policymakers may seek advice.

The rest of the paper proceeds as follows. First, we discuss the data used in the experiment. Then we describe and present results from the discrete choice experiment. Finally, we discuss the implications of our results.

[1]For example, in AidGrade's meta-analysis data, the median treatment effect of 36 conditional cash transfer programs on enrollment rates was 5.1 percentage points (AidGrade, 2016).

2 Data

We surveyed 190 policymakers at several World Bank and Inter-American Development Bank workshops, listed in Table 1 below. Of these, we obtained responses from 156, representing a very high response rate of 82%. We supplemented this sample by collecting data at the World Bank's headquarters in Washington, D.C., obtaining 16 additional responses. In total, across these two approaches we obtained responses from 172 policy professionals. The following subsections describe each sample in more detail.

2.1 World Bank Sample

We surveyed attendees at WB workshops organized in Athens (September 2019), Marrakesh (December 2019), Bangkok (July 2023), and Dar es Salaam (August 2023). Workshop attendees consisted of policymakers and policy practitioners.[2] Each workshop was approximately one week long and was meant to facilitate connections between government staff and researchers. Policymakers and policy practitioners were matched with researchers and worked together over the course of the workshop to design a prospective impact evaluation that could be used for their program

Workshops were attended by participants from around the world and we observed high response rates across workshops. We expect this is in part because of the data collection approach within the workshop format: surveys were conducted as part of the program agenda. The designated time slot was early in the workshops to mitigate the possibility of experimenter demand effects.

2.2 IDB Sample

Participants were also recruited from a workshop organized in May 2018 at the IDB headquarters in Washington, DC. Like the WB workshops, this workshop ran for approximately one week, was attended by policymakers and policy practitioners, and focused on impact evaluation methods. This workshop, however, did not include matching participants with researchers to design an impact evaluation as the World Bank workshops did.

[2] While academics and other researchers participated in these workshops, there were very few of them, and they will not be included in analysis.

Table 1: Response Rate at Workshops

Institution	Location	Year	Eligible Attendees	Surveyed	Response Rate
IDB	Washington, D.C.	2018	49	18 (18)	0.37 (0.37)
World Bank	Athens, Greece	2019	39	38	0.97
World Bank	Marrakesh, Morocco	2019	41	33	0.80
World Bank	Bangkok, Thailand	2023	30	26	0.87
World Bank	Dar es Salaam, Tanzania	2023	31	26	0.84
Total			190	156	0.82

The IDB rows include responses from the "pre" period and, in parentheses, the "post" period. The total is calculated using the total number of unique respondents across both rounds of the IDB survey. We exclude researchers from both the eligible and response counts, as too few attended to be considered. This excludes 12 researcher responses from the World Bank workshop in Athens and two from the World Bank workshop in Marrakesh. Overall, response rates were very high.

Participants were emailed the survey link by the workshop organizers before the start of the workshop and again after the workshop (for a second, identical survey). We focus on the survey responses collected before the start of the workshop, as these may be more representative of the typical preferences held by policymakers and policy practitioners.

2.3 World Bank Office Survey

We collected additional survey responses at the World Bank's main offices in Washington, D.C. to supplement our sample. We recruited participants by setting up a table outside the main cafeteria during lunch on August 2 - August 10, 2023 at which passers-by were invited to take the survey. This approach resulted in an additional 16 responses being collected.[3]

In total, our sample consists of 172 policy professionals, balanced between policymakers and policy professionals.

[3]This number excludes any respondents who did not meet our inclusion criteria, for example those working in IT, legal services, or administration.

Table 2: Attributes and Levels in the Discrete Choice Experiment

Attributes	Levels
Method	Experimental, Quasi-experimental, Observational
Location	Same country, Different country in the same region Different region
Impact	0, +5, +10 percentage points
Confidence Interval	+/−1, +/−10 percentage points
Recommended	Yes, No

This table shows the different attributes used in the discrete choice experiment and their levels.

3 Method

Participants were asked to repeatedly choose which of two programs they would prefer.[4] The programs were identified only as *Program A* or *Program B*, and each was associated with a study. These studies varied by method (experimental, quasi-experimental or observational); location (same country, different country in the same region, different country in a different region); the effect the study found (an increase in enrollment rates by 0, 5, or 10 percentage points); the precision of the estimate (a confidence interval of +/- 1 percentage point or +/- 10 percentage points); and whether a local expert recommended it. These attributes are summarized in Table 2. Respondents saw one block of six questions each at the World Bank workshops in Athens, Marrakesh, Bangkok and Dar es Salaam, two blocks at the IDB, and one block at the survey at the World Bank office. We used a fractional factorial design to select the variation of choice characteristics within blocks to optimize power. Results are analyzed using conditional logistic regression, clustering at the individual level.

The questions asked were hypothetical. We expect that this might tend to bias our results towards zero, since individuals would have less motivation to consider each question carefully. No incentives were provided for completing the survey, with the exception that at the World Bank headquarters survey, respondents were rewarded with a token gift of chocolate for completion. Since the study aims to elicit partici-

[4] E.g. "Now imagine that you need to provide a recommendation to a counterpart agency in your country on which of two programs to implement. A study was done on each program, with the results below. Please select which program you would recommend." Appendix Figure A1 shows an example of a choice scenario participants might have faced.

pants' unbiased beliefs and there are no clearly "correct" answers, we did not provide incentives that depended on participant responses. The surveys were anonymous to reduce the potential for experimenter demand effects.

4 Results

Table 3 presents results from the discrete choice experiment using a conditional logistic regression. Standard errors are clustered at the individual level. Policymakers preferred programs with larger estimated treatment effects or programs that came recommended by a local expert, as well as those that had an impact evaluation from their country (Column 2). Policy practitioners preferred programs with larger, more precisely estimated impact evaluation results as well as results from the same country as the target program and results from RCTs (Column 3). They also preferred programs that came recommended by a local expert.

Table 4 translates these results into estimates of how much policymakers and policy practitioners would be willing to pay, in terms of estimated impact, for programs with these different attributes. Policymakers would prefer a program recommended by a local expert even if it had an approximately 5.4 percentage point lower estimated impact on enrollment rates (Column 2), which is very large relative to the range of effects that programs typically have on enrollment rates. Results appear largely comparable across the World Bank and IDB pre-workshop samples. Table A1 tests whether policymakers and policy practitioners put statistically significantly different weights on different attributes. They largely do not, though policymakers put less weight on precise estimates.

To further explore the importance placed on precise estimates, we construct an indicator for whether the result shown was significant. While this variable was not explicitly shown to participants, it could be discerned from the provided estimated impact and confidence interval. When we include this variable in the regressions, the preference towards programs with studies having larger and more precisely-estimated effects goes away, suggesting it is driven by significance (Table A2). However, a limitation of these results is that the study was not set up to explicitly test this hypothesis and the levels of the attributes that were used to construct this variable only very crudely capture significance.

Finally, given that policy professionals appear to care about two forms of local

Table 3: Weighing Different Kinds of Evidence

	Pooled (1)	Policymaker (2)	Policy Practitioner (3)
Impact	1.062***	1.050***	1.075***
	(0.013)	(0.018)	(0.019)
Quasi-Experimental	1.074	0.975	1.184
	(0.114)	(0.143)	(0.183)
Experimental	1.355***	1.202	1.528**
	(0.153)	(0.179)	(0.263)
Different country, same region	1.070	1.033	1.134
	(0.104)	(0.142)	(0.160)
Same country	1.509***	1.364**	1.731***
	(0.132)	(0.173)	(0.209)
Recommended	1.262***	1.298**	1.234**
	(0.094)	(0.150)	(0.119)
Small C.I.	1.294***	1.071	1.565***
	(0.098)	(0.111)	(0.169)
Observations	912	431	481

This table reports the results of conditional logit regressions on which program was selected. Odds ratios are reported. "Impact" refers to the estimate of the effect associated with the program; "Quasi-Experimental" indicates whether the study associated with the program was quasi-experimental; "Experimental" indicates whether the study associated with the program was an RCT; "Different country, same region" indicates whether the study associated with the program was described as done in a different country in the same region; "Same country" indicates whether the study associated with the program was described as done in the same country; "Recommended" indicates whether the program was recommended by a local expert; "Small C.I." refers to the estimates having small confidence intervals. The omitted categories are "Observational", "Different region", and "Large C.I.". The number of observations represents the total number of choices made across individuals. Standard errors are provided in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table 4: Willingness to Pay (in Terms of Estimated Impact)

	Pooled (1)	Policymaker (2)	Policy Practitioner (3)
Quasi-Experimental	-1.187	0.530	-2.328
	(1.751)	(3.064)	(2.188)
Experimental	-5.032***	-3.794	-5.835**
	(2.055)	(3.202)	(2.728)
Different country, same region	-1.113	-0.664	-1.730
	(1.642)	(2.876)	(1.960)
Same country	-6.815***	-6.388**	-7.560***
	(1.906)	(3.121)	(2.478)
Recommended	-3.854***	-5.379*	-2.896**
	(1.423)	(3.008)	(1.418)
Small C.I.	-4.268***	-1.415	-6.166***
	(1.369)	(2.121)	(1.849)
Observations	912	431	481

This table reports the results of conditional logit regressions in terms of the implied willingness-to-pay for programs with certain attributes. "Impact" is implicitly included in each estimate. For example, in the pooled sample, policymakers would only be willing to accept a program that was not recommended over one that was if the program that was not recommended had a 5.4 percentage point higher estimated impact on enrollment rates. The number of observations represents the total number of choices made across individuals. Standard errors are provided in parentheses.

evidence - the recommendations from local experts and having impact evaluation evidence from the same setting - it is natural to wonder whether they view these as substitutes. The main results suggest this is not the case, since both variables are independently significant (Table 3). However, to further explore this issue, we interact whether a program was recommended by a local expert with whether impact evaluation results from the same country were displayed. Results in Table 5 suggest that policymakers do not put any more or less weight on advice from a local expert depending on whether an impact evaluation done in the same country is available, though it is possible that this is a function of limited power and the direction of the estimate for the interaction term is consistent with policymakers considering the two forms of local evidence to be substitutes.[5]

Research results serve as only one input into decision-making processes. While fully unpacking the role of different sources of information in policy decisions is beyond the scope of this short paper, we make a simple distinction between two commonly-used sources of evidence and consider how much weight policymakers place on them. Because of our experiment's simplicity, there are necessary caveats. In particular, policymakers in our setting choose between hypothetical programs and their decisions may differ under different circumstances. Further, the information we provide about these programs is necessarily limited in order to focus attention on the attributes of interest. Nonetheless, we hope that this short paper highlights policymakers' strong demand for local evidence.

5 Conclusion

We find that policymakers place considerable weight on programs that are recommended by local experts as well as those with impact evaluations from their own country. The estimated policymaker preferences can have startling implications: according to our willingness-to-pay estimates, if a policymaker were considering a program aimed at increasing enrollment rates they would be willing to accept a program

[5] An important caveat here is that merely having impact evaluation results from the same country does not mean those impact evaluation results were promising. The average impact evaluation result was positive (representing an increase in enrollment rates of 5 percentage points), but the weight policymakers place on programs being recommended by local experts could in principle stem from those instances in which an impact evaluation showed no effect. The interaction term remains insignificant when all attribute levels are included, with the same caveats (results available upon request).

Table 5: Having an Impact Evaluation from the Same Country is not Seen as a Substitute for Local Expertise

	Pooled (1)	Policymaker (2)	Policy Practitioner (3)
Impact	1.061***	1.052***	1.069***
	(0.013)	(0.017)	(0.018)
Same country	1.537**	1.553*	1.533
	(0.285)	(0.401)	(0.409)
Recommended	1.295**	1.440**	1.179
	(0.165)	(0.261)	(0.212)
Same country * Recommended	0.863	0.748	0.976
	(0.289)	(0.334)	(0.482)
Observations	912	431	481

This table reports the results of conditional logit regressions on which program was selected. Odds ratios are reported. The significance of the "Recommended" dummy and the insignificant interaction between "Same country" and "Recommended" in Column 2 suggests that policymakers do not put more or less weight on advice from a local expert depending on whether another form of local evidence - an impact evaluation done in the same country - is available. The number of observations represents the total number of choices made across individuals. Standard errors are provided in parentheses. * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

that had been shown to have a 5.4 percentage point lower impact if the program came recommended by a local expert. They would also be willing to accept a program shown to have a 6.4 percentage point lower impact if it had been evaluated in their own country. These trade-offs can be larger than the effect of the typical program. These results suggest that unless research is seen by policymakers as valid in their target setting, policymakers are likely to choose alternative programs that may have lower estimated treatment effects but be from a better-fitting context.

Given that individuals like those in our sample approve and implement many development programs, evidence on how they value the results that research provide can help us understand how to design studies such that they better feed into the evidence-to-policy pipeline. Our results imply that researchers aiming to improve policy should try to design studies to approximate the target context as closely as possible to maximize the chance that their results are taken up by policymakers.

References

AidGrade (2016). AidGrade Impact Evaluation Data, Version 1.3.

Bessone, P., G. Rao, F. Schilbach, H. Schofield, and M. Toma (2021). The Economic Consequences of Increasing Sleep Among the Urban Poor. *Quarterly Journal of Economics 136*(3), 1887–1941.

Hanna, R., S. Mullainathan, and J. Schwartzstein (2014). Learning through noticing: Theory and evidence from a field experiment. *The Quarterly Journal of Economics 129*(3), 1311–1353.

Hjort, J., D. Moreira, G. Rao, and J. F. Santini (2019). How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities. *NBER Working Paper*.

Iacovone, L., D. McKenzie, and R. Meager (2023). Bayesian impact evaluation with informative priors: An application to a colombian management and export improvement program. *World Bank Policy Research Working Papers*.

Leight, J. (2022). Patterns of randomization, geographic concentration and funding in the recent development economics literature. *Working Paper*.

Meisch, S. P., S. Bremer, M. T. Young, and S. O. Funtowicz (2022). Extended peer communities: Appraising the contributions of tacit knowledges in climate change decision-making. *Futures 135*, 102868.

Milkman, K. L., L. Gandhi, M. S. Patel, H. N. Graci, D. M. Gromet, H. Ho, J. S. Kay, T. W. Lee, J. Rothschild, J. E. Bogard, and et al. (2022). A 680,000-person megastudy of nudges to encourage vaccination in pharmacies. *Proceedings of the National Academy of Sciences*.

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National academy of Sciences 111*(20), 7176–7184.

Nellis, G., T. Dunning, G. Grossman, M. Humphreys, S. D. Hyde, C. McIntosh, and C. Reardon (2019). *Information, Accountability, and Cumulative Learning*, Chapter Learning about Cumulative Learning: An Experiment with Policy Practitioners. Cambridge University Press.

Podgórski, D. (2010). The use of tacit knowledge in occupational safety and health management systems. *International Journal of Occupational Safety and Ergonomics 16*(3), 283–310.

Polanyi, M. (2009). The tacit dimension. In *Knowledge in organisations*, pp. 135–146. Routledge.

Pritchett, L. and J. Sandefur (2015). Learning from experiments when context matters. *American Economic Review 105*(5), 471–475.

Rogger, D. and R. Somani (2023). Hierarchy and information. *Journal of Public Economics 219*, 104823.

Toma, M. and E. Bell (2022). Understanding and improving policymakers' sensitivity to program impact.

Vivalt, E. (2020). How much can we generalize from impact evaluations? *Journal of the European Economic Association 18*(6), 3045–3089.

Vivalt, E., A. Coville, and S. KC (2023). Seeking Evidence: Which Studies Do Policymakers Prefer? *Working Paper*.

Appendices

A Additional Tables and Figures

Table A1: Comparing Policymakers and Policy Practitioners

	Pooled (1)	World Bank (2)	IDB (3)
Impact	1.075***	1.062***	1.134***
	(0.019)	(0.021)	(0.046)
Quasi-Experimental	1.184	1.260	0.913
	(0.183)	(0.223)	(0.247)
Experimental	1.528**	1.558**	1.461
	(0.262)	(0.304)	(0.607)
Different country in the same region	1.134	1.160	1.061
	(0.160)	(0.204)	(0.223)
Same country	1.731***	1.754***	1.751***
	(0.208)	(0.250)	(0.362)
Recommended	1.234**	1.202*	1.397
	(0.118)	(0.131)	(0.300)
Small C.I.	1.565***	1.573***	1.581*
	(0.169)	(0.195)	(0.394)
Policymaker * Impact	0.976	0.982	0.953
	(0.024)	(0.026)	(0.060)
Policymaker * Quasi-Experimental	0.823	0.737	1.080
	(0.175)	(0.172)	(0.577)
Policymaker * Experimental	0.787	0.773	0.722
	(0.179)	(0.197)	(0.432)
Policymaker * Different country,	0.911	0.972	0.685
same region	(0.179)	(0.226)	(0.260)
Policymaker * Same country	0.788	0.765	0.868
	(0.138)	(0.154)	(0.334)
Policymaker * Small C.I.	0.685**	0.655***	0.759
	(0.102)	(0.104)	(0.295)
Policymaker * Recommended	1.052	1.006	1.260
	(0.157)	(0.172)	(0.393)
Observations	912	705	207

This table reports the results of conditional logit regressions on which program was selected. Odds ratios are reported. Interaction terms are used to test whether policymakers weigh different attributes of studies differently. “Impact” refers to the estimate of the effect associated with the program; “Quasi-Experimental” indicates whether the study associated with the program was quasi-experimental; “Experimental” indicates whether the study associated with the program was an RCT; “Different country, same region” indicates whether the study associated with the program was described as done in a different country in the same region; “Same country” indicates whether the study associated with the program was described as done in the same country; “Recommended” indicates whether the program was recommended by a local expert; “Small C.I.” refers to the estimates having small confidence intervals. The omitted categories are “Observational”, “Different region”, and “Large C.I.”, as well as the equivalent categories interacted with the policymaker dummy. The number of observations represents the total number of choices made across individuals. Standard errors are provided in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Table A2: Significance is Important to Policymakers and Policy Practitioners

	Pooled	Policymaker	Policy Practitioner
	(1)	(2)	(3)
Impact	1.022	1.013	1.033
	(0.017)	(0.023)	(0.025)
Quasi-Experimental	1.141	1.041	1.250
	(0.122)	(0.150)	(0.198)
Experimental	1.400***	1.236	1.582***
	(0.158)	(0.181)	(0.273)
Different country, same region	1.109	1.076	1.170
	(0.109)	(0.145)	(0.170)
Same country	1.510***	1.381**	1.712***
	(0.132)	(0.177)	(0.205)
Recommended	1.304***	1.337**	1.280**
	(0.100)	(0.160)	(0.125)
Small C.I.	0.789	0.671	0.940
	(0.141)	(0.167)	(0.241)
Significant	2.130***	2.046*	2.169**
	(0.554)	(0.773)	(0.789)
Observations	912	431	481

This table reports the results of conditional logit regressions on which program was selected. Odds ratios are reported. "Impact" refers to the estimate of the effect associated with the program; "Quasi-Experimental" indicates whether the study associated with the program was quasi-experimental; "Experimental" indicates whether the study associated with the program was an RCT; "Different country, same region" indicates whether the study associated with the program was described as done in a different country in the same region; "Same country" indicates whether the study associated with the program was described as done in the same country; "Recommended" indicates whether the program was recommended by a local expert; "Small C.I." refers to the estimates having small confidence intervals; "Significant" indicates whether a result might be perceived as significant according to its point estimate and confidence interval. The omitted categories are "Observational", "Different region", "Large C.I." and "Insignificant". The number of observations represents the total number of choices made across individuals. Standard errors are provided in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

Figure A1: Example of a Choice Scenario

Now imagine that you need to provide a recommendation to a counterpart agency in your country on which of two programs to implement. A study was done on each program, with the results below. Please select which program you would recommend.

	Study on Program A	*Study on Program B*
Method	Observational	Quasi-experimental
Location	A country in a different region	Same country
Impact on enrollment rates, with margin of error (95% confidence interval)	0 percentage point, +/-10 percentage points	+10 percentage points, +/-1 percentage point

A local expert tells you that they believe Program B would perform better in your context.

Which program do you recommend?

Program A

Program B