

Fischer, Mira; Grewenig, Elisabeth; Lergetporer, Philipp; Werner, Katharina; Zeidler, Helen

Article — Published Version

The E-word – On the public acceptance of experiments

Economics Letters

Provided in Cooperation with:

WZB Berlin Social Science Center

Suggested Citation: Fischer, Mira; Grewenig, Elisabeth; Lergetporer, Philipp; Werner, Katharina; Zeidler, Helen (2024) : The E-word – On the public acceptance of experiments, Economics Letters, ISSN 1873-7374, Elsevier, Amsterdam, Vol. 235, pp. 1-3, <https://doi.org/10.1016/j.econlet.2024.111558>

This Version is available at:

<https://hdl.handle.net/10419/323980>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>



The E-word – On the public acceptance of experiments[☆]

Mira Fischer^a, Elisabeth Grewenig^b, Philipp Lergetporer^{c,*}, Katharina Werner^d, Helen Zeidler^e

^a WZB Berlin, Germany

^b Kreditanstalt für Wiederaufbau (KfW), Germany

^c Technical University of Munich, Campus Heilbronn, ifo Institute, CESifo and IZA, Germany

^d ifo Institute at the University of Munich, Germany

^e Technical University of Munich, Campus Heilbronn, Germany

ARTICLE INFO

January 16, 2024

JEL classification:

I28

H40

C93

Keywords:

Experiment aversion

Policy experimentation

Education

ABSTRACT

Randomized experiments are often viewed as the “gold standard” of scientific evidence, but people’s skepticism towards experiments has compromised their viability in the past. We study preferences for experimental policy evaluations in a representative survey in Germany ($N > 1,900$). We find that a majority of 75 % supports the idea of small-scale evaluations of policies before enacting them at a large scale. Experimentally varying whether the evaluations are explicitly described as “experiments” has a precisely estimated overall zero effect on public support. Our results indicate political leeway for experimental policy evaluation, a practice that is still uncommon in Germany.

1. Introduction

Randomized experiments are often referred to as the “gold standard” of scientific evidence, and their use in natural contexts has markedly increased over the past two decades (Baldassarri and Abascal, 2017). Field experiments have transformed from being primarily small-scale proof-of-concept studies (Grose, 2014) into a broader tool for program evaluation in various fields, and further, into a comprehensive approach for governance and policymaking (Huitema et al., 2018). The experimental turn in policymaking is highlighted by the growing presence of government advisors with expertise in experimental social science and by the OECD’s advocacy for policy experiments (OECD, 2019). Education is a particularly important field in which randomized controlled trials are proliferating and helping to improve policy (Sadoff, 2014).

At the same time, backlash from political decision-makers, bureaucrats, study participants, the public at large, and other stakeholders can compromise the viability of randomized field experiments (e.g.,

Heckman and Smith, 1995; Heffetz and List, 2021). A case in point is Angrist and Lavy (2002), in which the authors report that an experiment offering cash incentives for students was suspended after “extensive and mostly critical media coverage”. Other examples for experiments facing strong public condemnation after their implementation include the Facebook newsfeed experiment (Goel, 2014) or the matching score experiment of the dating platform OKCupid (Hern, 2014). While the public’s acceptance of experiments is crucial for their feasibility, little systematic evidence exists on the extent and determinants of people’s support for experimental policy evaluation.

We investigate the public’s preferences for reform evaluation, and test the hypothesis that explicitly describing an evaluation as an “experiment” triggers public backlash. Negative reactions to the word “experiment” might be due to several reasons. It may make citizens think of past unethical or even criminal studies that have been referred to as “experiments” (e.g., the crimes against humanity committed by Nazi doctors during World War II)¹, it might also trigger concerns about

[☆] We are most grateful to the editor Audra Bowlus, an anonymous referee, Dorothea Kuebler and Ludger Woessmann for their support and advice, and to Franziska Kugler for her help in preparing the survey. Financial support by the Leibniz Competition (SAW-2014-ifo-2) and the German Science Foundation (CRC TRR 190) is gratefully acknowledged.

* Corresponding author.

E-mail address: philipp.lergetporer@tum.de (P. Lergetporer).

¹ These horrific crimes led to the creation of the Nuremberg Code of 1947, a code of research ethics for medical experimentation with human subjects (see also List, 2008).

policy uncertainty, as exemplified by the successful 1957 federal-election campaign slogan “*Keine Experimente!*” (“No Experiments!”) of Germany’s chancellor Konrad Adenauer.² Finally, the word “experiment” might prime reportedly unpopular features of field experiments – such as the use of randomization, denial of treatment to control-group members (e.g., Heckman and Smith, 1995), or lack of informed consent (e.g., List, 2008). Anecdotal evidence suggests that experimental economists often avoid the word “experiment” when communicating their research because they fear that using the word may yield backlash. If merely avoiding the word “experiment” can foster the political feasibility of field experiments, altered communication strategies could make conducting field experiments much easier and mitigate reluctance to use them.

We conduct a survey experiment among a representative sample of the German voting-age population ($N > 1900$) in which respondents are randomly assigned to one of two versions of a question on preferences for reform evaluation. Focusing on education policy, the baseline version of the question describes the evaluation process without explicitly mentioning the word “experiment”. In the treatment group, we used the exact same wording as in the control group with the sole exception that the description of the evaluation process additionally includes the phrase “with experiments”.

We find that a clear majority in the control group supports the idea of evaluating education reforms before rolling them out at a large scale: 75 % are in favour of the proposal and only 14 % oppose it (the remaining 11 % are indifferent). Using the word “experiment” to describe reform evaluations has a precisely estimated zero causal effect on overall public support for education policy evaluation. In additional analyses, we show that treatment effects are homogeneous across sociodemographic subgroups, and that results are unlikely to be driven by respondents’ inattention.

Our study contributes to the emerging experimental literature on preferences for (policy) experimentation among the public (e.g., Meyer et al., 2019; Mislavsky et al., 2019) and policymakers (e.g., Dur et al., 2023; Vivalt et al., 2023). We complement this strand of research by providing first evidence on how the use of the word “experiment” affects public preferences for experimental reform evaluation.

2. Experimental setup

The experiment was embedded in the 2017 wave of the *ifo Education Survey*, an annual opinion survey on education policy in Germany, a country where experimental policy evaluation is rarely conducted (see Appendix B for details about the survey and the institutional setting).³ Our goal is to investigate whether using the word “experiment” to describe the evaluation of educational reforms affects public support for reform evaluation. Therefore, we randomly assigned respondents to one of two versions of a question that elicits public preferences for education reform evaluation. The control-group version of the question was worded as follows: “*Do you support or oppose that the effects of reforms in the education system, just like new medicine, should initially be tested on a small scale before they are implemented nationwide?*” In contrast, the treatment-group question reads as follows: “*Do you support or oppose that the effects of reforms in the education system, just like new medicine, should initially be tested using experiments on a small scale before they are implemented nationwide?*” Note that the question wording is identical across experimental groups, with the sole exception being that in the treatment group the words “with experiments” were added. Respondents were asked to select one of the following five answer categories: strongly support,

somewhat support, neither support nor oppose, somewhat oppose, strongly oppose (see Appendix Fig. A1 for screenshots). Using a fair coin, we randomly assigned 949 respondents to the treatment group and 1016 respondents to the control group. Appendix Table A1 presents sample characteristics and shows that randomization worked as intended. We estimate treatment effects using simple OLS models (see Appendix B for the empirical model).

3. Results

Table 1 depicts our main results. Odd-numbered columns present estimates without controls, even-numbered columns include our set of sociodemographic control variables. A 75 % majority of respondents in the control group supports the evaluation of education reform (see control mean), only 14 % oppose it. The remainder is neutral. This widespread support echoes the majority backing for policy experimentation in other areas, for example, among Dutch voters as documented by Dur et al. (2023).

The small and statistically insignificant coefficients on the treatment indicator in Table 1 show that using the phrase “experiment” to describe reform evaluations does not affect average support for the evaluation of educational reforms. Note that the estimated effects are very small and that we are powered to detect treatment effects of 6 percentage points in columns 1 and 2, and 5 percentage points in columns 3 and 4.⁴ In Appendix Table A2 we show that the treatment has small and insignificant effects on each of the five answer categories. Appendix Table A3 (column 1) shows how support for reform evaluation with experiments varies with respondents’ characteristics. In sum, the vast majority of Germans supports the evaluation of educational reforms, even if this evaluation is clearly labelled with the “E-word”.

Next we explore variations in treatment effects across respondents’ sociodemographic characteristics in an exploratory heterogeneity analysis. A priori, one may expect that, for instance, reactions to the word “experiment” are more negative among parents with school-aged children or less risk-tolerant individuals. However, treatment effects do not differ across sociodemographic subgroups defined by, for

Table 1
Effects of using the word “experiment” on preferences for reform evaluation.

	Support for reform evaluation		Opposition reform evaluation	
	(1)	(2)	(3)	(4)
„Experiment“ treatment	−0.009 (0.022)	−0.015 (0.023)	0.019 (0.018)	0.021 (0.018)
Covariates	No	Yes	No	Yes
Control mean	0.751		0.140	
Observations	1957	1902	1957	1902
R ²	0.000	0.022	0.001	0.024

Notes: OLS regressions. “Experiment” treatment: 1 = word “experiment” is included in the question text, 0 otherwise. Dependent variable: Columns 1–2: Dummy variables 1 = “strongly support” or “somewhat support” reform evaluation, 0 otherwise; columns 3–4: Dummy variables 1 = “strongly oppose” or “somewhat oppose” reform evaluation, 0 otherwise. Residual category: “neither support nor oppose.” Control mean: mean of the outcome variable in the control group. Data source: ifo Education Survey 2017. Regressions weighted by survey weights. Robust standard errors in parentheses. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

example, parenthood, educational background, employment in

² The slogan was used by the Christian Democratic Union (CDU) and referred to the risk that the Social Democratic Party (SPD) would leave the NATO in case of electoral victory.

³ The data from the ifo Education Survey are available for scientific use (Freundt et al., 2022). See Appendix B for details.

⁴ To compute minimal detectable effect sizes with 80% power and $\alpha=0.05$, we follow Haushofer and Shapiro (2016) and multiply standard errors by 2.8.

education or risk tolerance (column 2 of Appendix Table A3). The only exception is gender, with females reacting more negatively to the word “experiment” than males.⁵ Importantly, we find no heterogeneity by response time or educational background, indicating that the overall zero effect is not due to respondents’ inattention or misunderstanding.

4. Conclusion

In our representative survey experiment, we find majority support for scientific reform evaluation, irrespective of whether it is termed as an “experiment”. While there is widespread publication bias against null results (Chopra et al., 2023), we consider reporting this zero effect important as it provides the first causal evidence on whether using the “E-word” causes public backlash.

Our results shed light on the effect of the terminology used to describe scientific policy evaluation in the context of education policy. We consider the educational sector particularly interesting, since the group that may be impacted by experimentation are children, who are particularly vulnerable. At the same time, our experimental setup has certain limitations: First, it is not clear whether our results on education policy extend to other policy areas. Second, we did not measure which exact elements of policy evaluation respondents have in mind when answering the control-group or treatment-group version of the question. Thus, our study design does not facilitate investigating which elements of (experimental) policy evaluation, like randomization, influence public attitudes.⁶ Third, and relatedly, the description of the policy evaluation in both experimental groups as “just like new medicine” may also prime control-group respondents to think about experimental policy evaluation, limiting the scope for possible treatment effects of using the “E-word” explicitly. We think that conducting additional experiments to study the replicability, generality, and robustness of our experimental findings is a valuable avenue for future research.

Data availability

The data is available upon request.

References

- Angrist, J.D., Lavy, V., 2002. The effect of high school matriculation awards: evidence from randomized trials. NBER Working Pap. 9389.
- Baldassarri, D., Abascal, M., 2017. Field experiments across the social sciences. *Annu. Rev. Sociol.* 43, 41–73.
- Chopra, F., Haaland, I., Roth, C., Stegmann, A., 2023. The null result penalty. *Econ. J.*, forthcoming.
- Dur, R., Non, A., Prottung, P., Ricci, B., 2023. Who’s afraid of policy experiments? Tinbergen Institute Discussion Papers, pp. 23–027. /V.
- Freundl, V., Grewenig, E., Kugler, F., Lergetporer, P., Schöler, R., Wedel, K., Werner, K., Wirth, O., Woessmann, L., 2022. The ifo education survey 2014–2021: a new dataset on public preferences for education policy in Germany. *J. Econ. Stat.* ahead of print.
- Goel, V., 2014. Facebook tinkers with users’ emotions in news feed experiment, stirring outcry. *New York Times*, 29/07/2014.
- Grose, C.R., 2014. Field experimental work on political institutions. *Ann. Rev. Polit. Sci.* 17, 355–370.
- Haushofer, J., Shapiro, J., 2016. The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya. *Q. J. Econ.* 131, 1973–2042.
- Heckman, J.J., Smith, J.A., 1995. Assessing the case for social experiments. *J. Econ. Perspect.* 9, 85–110.
- Heffetz, O., List, J., 2021. Who’s afraid of evidence-based policymaking? Project Syndicate.
- Hern, A. (2014). OKCupid: we experiment on users. Everyone does, the guardian 07/24/2014, <https://www.theguardian.com/technology/2014/jul/29/okcupid-experiment-human-beings-dating>, [accessed 27 September 2023].
- Huitema, D., Jordan, A., Munaretto, S., Hildén, M., 2018. Policy experimentation: core concepts, political dynamics. *Gov. Impacts* 51 (2), 143–159. *Policy Sciences*.
- List, J.A., 2008. Informed consent in social science. *Science* 322 (5886), 672.
- Meyer, M.N., Heck, P.R., Holtzman, G.S., Anderson, S.M., Cai, W., Watts, D.J., Chabris, C. F., 2019. Objecting to experiments that compare two unobjectionable policies or treatments. *PNAS* 116 (22), 10723–10728.
- Mislavsky, R., Dietvorst, B., Simonsohn, U., 2019. The minimum mean paradox: a mechanical explanation for apparent experiment aversion. *PNAS* 116 (48), 23883–23884.
- OECD, 2019. Tools and Ethics for Applied Behavioural Insights: The BASIC Toolkit. OECD Publishing, Paris.
- Sadoff, S., 2014. The role of experimentation in education policy. *Oxf. Rev. Econ. Policy* 30 (4), 597–620.
- Vivalt, E., Coville, A., Sampada, K.C., 2023. Weighing the evidence: which studies count? Working Pap.

⁵ We also find marginally significant heterogeneities by respondents’ country of birth and political leaning. However, unlike the heterogeneity by gender, these heterogeneities become insignificant after adjusting for multiple-hypothesis testing in our heterogeneity analysis (see Appendix Table A3).

⁶ The presence or absence of randomization is one particularly interesting element, since studies examining aversion to randomization yield conflicting results (Meyer et al., 2019; Mislavsky et al., 2019).