# Pre-Analysis Plan

## Performance Gaps in Education: An Experimental Investigation of Public Perception and Policy Support

## Content

# RCT Registration

**Trial title**: Performance Gaps in Education: An Experimental Investigation of Public Perception and Policy Support

**Status:** In development

**Country:** Germany

**Primary investigator**: Fabian Mierisch (Catholic University Eichstaett-Ingolstadt)

**Other primary investigator(s)**: -

**Key words**: Education, Inequality

**Additional Keywords**: Experiment, Education, Inequality

**JEL:** I24, C93, J15

**Abstract (will be publicly available on the web page):**

There are large performance gaps in standardized test scores based on gender, migration background, and socioeconomic status. However, these gaps are the result of an interplay of individual choices and institutional inequality. Therefore, it is unclear if the public sees these performance gaps as a problem, and if there is a need to implement policies to compensate for differences. To investigate these questions, I conduct an information provision experiment using a representative sample of Germans (N~4,000). In the experiment, I provide participants with information regarding performance gaps in education. I then assess if survey participants perceive performance gaps as problematic and their views on appropriate measures to address these disparities. The findings potentially have implications for educational policies and interventions aimed at reducing inequalities and promoting equal opportunities in the German education system.

**Trial Start date:** 06.05.2019

**Intervention Start Date**: 06.05.2019

**Intervention End Date**: 24.05.2019

**Trial End Date:** 31.12.2024

**Sponsors:** ifo Institute, LMU-ifo Economics & Business Data Center

**Partners:** None

# Experimental Details:

## Intervention (public):

I utilize a Germany-wide representative survey (Freundl et al., 2022) to conduct an information provision experiment following the approach proposed by Haaland et al. (2023). Within the survey, I employ a random assignment strategy to provide participants with information regarding performance gaps in standardized tests. Specifically, I present information on the disparities observed between genders, children with and without a migrant background, and children from parents with different socioeconomic statuses. Subsequently, I examine the impact of this information on participants' perception of the problematic nature of these gaps and their support for measures aimed at addressing performance differences between girls and boys. Additionally, I explore the treatment effects on participants' perception of public spending in the field of education through an obfuscated follow-up.

References:

Vera Freundl, Elisabeth Grewenig, Franziska Kugler, Philipp Lergetporer, Ruth Schüler, Katharina Werner, Katharina Wedel, Olivia Wirth, Ludger Wößmann, „The ifo Education Survey 2014–2021: A New Dataset on Public Preferences for Education Policy in Germany“, Jahrbücher für Nationalökonomie und Statistik, 2022.

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. Journal of Economic Literature, 61(1), 3-40.

## Intervention (Hidden):

I employ a Germany-wide representative survey (Freundl et al., 2022) to conduct an information provision experiment based on the methodology proposed by Haaland et al. (2023). The intervention begins by eliciting participants' prior beliefs about the relative performance of different groups in a standardized test via the following questions:

A test regularly examines the mathematics performance of 4th grade elementary school students in Germany (TIMSS = Trends in International Mathematics and Science Study).

How do you think each of the following groups of children compare on the test?

i) … What do you estimate is the case for girls and boys?
ii) … What do you estimate, how is that for children from different social backgrounds?
iii) … What do you estimate, how is that with children without and with a migration background?

Responses are recorded on an eleven-point Likert scale, with one group positioned on one end and the other group on the opposite end (e.g., "1 - girls are much better," "3 - girls are somewhat better," "6 - both equally good," "9 - boys are somewhat better," "11 - boys are much better"). To counteract ordering effects, the Likert scale is inverted for half of the participants. Additionally, participants indicate their level of confidence in the estimated response on a seven-point Likert scale ranging from "1 - very unconfident" to "7 - very confident" after each prior belief question.

Next, participants are randomized into four groups: two uninformed control groups and two treatment groups. The treatment groups receive information stating that boys, on average, perform 16 points better in the mathematics section of the standardized test or that girls, on average, perform 21 points better in the reading section of the standardized test. Following the provision of this information, participants are asked to express their support or opposition to the introduction of gender-specific support to compensate for performance differences between girls and boys. Responses are collected on a five-point Likert scale ranging from "1 - strongly favor" to "5 - strongly oppose."

### Control group 1:

Do you favor or oppose introducing gender-specific support to compensate for differences in mathematics performance between girls and boys?

### Treatment group 1:

The PISA test has shown that boys score on average 16 points higher than girls in the mathematics test.

Do you favor or oppose introducing gender-specific support to compensate for differences in mathematics performance between girls and boys?

### Control group 2:

Do you favor or oppose introducing gender-specific support to compensate for differences in reading performance between girls and boys?

### Treatment group 2:

The PISA test has shown that girls score on average 21 points higher than boys in the reading test.

Do you favor or oppose introducing gender-specific support to compensate for differences in reading performance between girls and boys?

I then randomize participants from the uninformed control groups into two additional groups. One uninformed control group and one group that receives information on average performance gaps between girls and boys (boys 3 points better), of children with and without migration background (children without a migration background 29 points better), and for children from parents with higher or lower socioeconomic status (children from parents with a goof background 40 points better). Participants then state whether they perceive the disparity to be a problem on a four point Likert scale (from “1 - not a problem at all", to “4 - a very serious problem").

### Control group:

What do you think, are unequal opportunities for children from the following groups in the German education system a serious problem?

Unequal opportunities for

i) … children from good or difficult social backgrounds.
ii) … boys and girls.
iii) … children with and without a migration background.

### Treatment group:

What do you think, are unequal opportunities for children from the following groups in the German education system a serious problem?

Unequal opportunities for

i) … children from good or difficult social backgrounds (those from good backgrounds score 40 points better).

ii) … boys and girls (boys are 3 points better).

iii) … children with and without a migration background (those without a migration background are 29 points better).

In an obfuscated follow-up, I also investigate treatment effects on the perception of public spending in education. Participants take the follow-up survey about 14 days after the initial survey. In the follow-up survey, I ask participants:

i) What do you guess, how much is spent on average each year per student on public general schools in Germany? [Answer as an integer in Euros.]

ii) In your opinion, should public spending for schools in Germany increase, decrease, or stay the same? [Answer on a five point Likert scale from “1 - greatly increase", to "5 - greatly decrease".]

References:

Vera Freundl, Elisabeth Grewenig, Franziska Kugler, Philipp Lergetporer, Ruth Schüler, Katharina Werner, Katharina Wedel, Olivia Wirth, Ludger Wößmann, „The ifo Education Survey 2014–2021: A New Dataset on Public Preferences for Education Policy in Germany“, Jahrbücher für Nationalökonomie und Statistik, 2022.

Haaland, I., Roth, C., & Wohlfart, J. (2023). Designing information provision experiments. Journal of Economic Literature, 61(1), 3-40.

## Primary Outcomes:

Policy Preference

First, I assess preferences for introducing gender-specific support to compensate for differences in. I measure preferences on a five point Likert scale from "1 – strongly favor" to "5 – strongly oppose".

Problem Perception

Second, I ask participants if they perceive unequal opportunities for children [from good or difficult social backgrounds/ with different genders/ with and without migration background] in the German education system as a problem. I measure this perception on a four point Likert scale from “1 - not a problem at all", to “4 - a very serious problem"

Follow-up Survey

Third, in a follow-up survey I measure the perception of governmental spending. I ask participants how much is spent on average each year per student on public general schools in Germany via an open entry field. Then I ask participants whether public spending on education should increase, decrease, or stay the same on a five point Likert scale from “1 - greatly increase", to "5 - greatly decrease".

## Secondary Outcomes:

I will conduct heterogeneity analyses by respondents' stated priors about the relative performance of the groups, by their certainty in their responses, and by demographics, such as e.g. migration background, parent status, and education background.

## Experimental Design (Public)

I randomize participants in four groups. The groups are two uninformed control groups and two treatment groups that receive information on performance gaps between girls and boys in a standardized test. Participants then have to answer a question whether they are in favor or oppose introducing gender-specific support to compensate for differences in mathematics/reading performance between girls and boys on a five point Likert scale. I then only randomize participants of the uninformed control group into two additional groups. One uninformed control group and one group that receives information on average performance gaps between girls and boys, of children with and without migration background, and for children from parents with higher or lower socioeconomic status. Participants then state whether they perceive the disparity to be a problem on a four point Likert scale. In an obfuscated follow-up, I also investigate treatment effects on the perception of public spending in education.

**Randomization Method** (e.g. public lottery, randomization done in office by a computer, coin flip): Randomization by survey software

**Randomization Unit**: Participant level

**Was the treatment clustered?** No

**Planned number of clusters:** -

**Sample size:** 4'000

**Power Calculation/ Minimum Detectable Effect Size for Main Outcomes:** With a sample of 4'000 respondents, four treatment arms and 80% power at the 5% significance level, the minimum detectable effect size is at around 0.052 of a difference in the mean scale points between two treatment conditions.

**IRB approval:** -

# Pre-Analysis Plan

## Empirical Strategy

I will estimate treatment effects by regressing a specific outcome of interest (i.e., policy preference, discrimination source) on randomized treatment indicators using OLS models. In order to increase precision (and to account for small imbalances between treatment and control groups), we will include a vector of control variables in our main specification.

Our main specification will be the following:

$$Y_i = \alpha + \beta_1 Treatment_i + \mu X_i + \varepsilon_i$$

$Y_i$: Outcome of interest of survey participant $i$

$Treatment_i$: Dummy equal to 1 if respondent received information treatment, 0 otherwise

$X_i$: Vector of control variables for respondent characteristics

$\varepsilon_i$: Idiosyncratic error term

With my randomized research design, the causal treatment effects of information provision on the respective outcome can be derived from $\beta_1$. Controls include gender, age, education, parental status, and federal state of residence of a survey participant. To avoid losing observations due to missing information, we will impute missing values in control variables and add imputation dummies to our regressions.

## Main Analyses

In the initial stage of analysis, I will estimate the impact of the treatment information on the outcomes of interest. These outcomes include support for gender-specific support to compensate for differences in mathematics/reading, the perception of performance gaps being problematic, the perception of public spending on education, and the need to increase public spending on education. To assess the treatment effects, I will measure the extent to which individuals agree with each outcome using Likert scales of varying response options (please refer to the description of primary outcomes for more detailed information regarding the Likert scales employed in this study). By quantifying the treatment effects on these outcome measures, I aim to evaluate the influence of the information provision on participants' attitudes and perceptions.

## Further Analyses

Furthermore, I intend to explore heterogeneities in treatment effects by considering several factors. Firstly, I will examine treatment effect heterogeneity based on respondents' stated priors about the relative performance of the groups. Through analyzing variations in treatment effects among participants with different initial beliefs, I aim to gain insights into the influence of preconceived notions on attitude change.

Moreover, I will investigate treatment effect heterogeneity based on respondents' level of certainty in their prior beliefs. Participants who express higher confidence in their initial responses may exhibit different patterns of attitude change compared to those with lower confidence levels.

I will also consider demographic characteristics to explore treatment effect heterogeneity. Variables such as migration background, parental status, and educational background will undergo analysis to determine whether these factors influence the impact of information provision on participants' perceptions and attitudes.

Lastly, I will conduct a subgroup analysis to compare treatment effects between participants who passed an attention check and those who did not. This analysis aims to assess the robustness of treatment effects among participants who demonstrated sufficient attentiveness during the survey.

By examining these various sources of heterogeneity, I aim to provide a nuanced understanding of the differential effects of information provision on participants' attitudes and perceptions.