Meritocracy, redistributive preferences and school political socialization

1. What is the hypothesis to be investigated?

Provide a brief description of the relevant theory and formulate the hypothesis as precisely as possible.

Increasing levels of economic inequality and income concentration have driven research on preferences for economic redistribution in recent years (Becker, 2021; Rueda & Stegmueller, 2019), which are understood as those beliefs about the need for economic transfers and social services for those who have less, generally through the State. However, most of the research on redistributive preferences has been carried out in the adult population, leaving aside the study of the factors associated with these preferences at earlier ages as well as their formation. Using data from 8th-grade students in Chile, in this research we are interested in delving into the possible relationships between the perception of meritocracy and preferences for redistribution at school age. Since meritocracy is conceived as a system where rewards are distributed according to individual effort and talent (Young, 1958), the central hypothesis of this research is that **those who perceive that there is more meritocracy will tend to show lower preferences for redistribution in school and society, as there are social conditions that allow individual achievements** (H_1) (Batruch et al, 2021).

Redistributive preferences have generally been investigated through public opinion surveys from different disciplines and various countries. There are different ways of operationalizing redistributive preferences, and for the case of this research we will use two of them. The first has to do with the preference for reducing economic gaps between rich and poor, and the second has to do with the justification (or not) for access to social benefits based on individual income, such as education, health, and pensions.

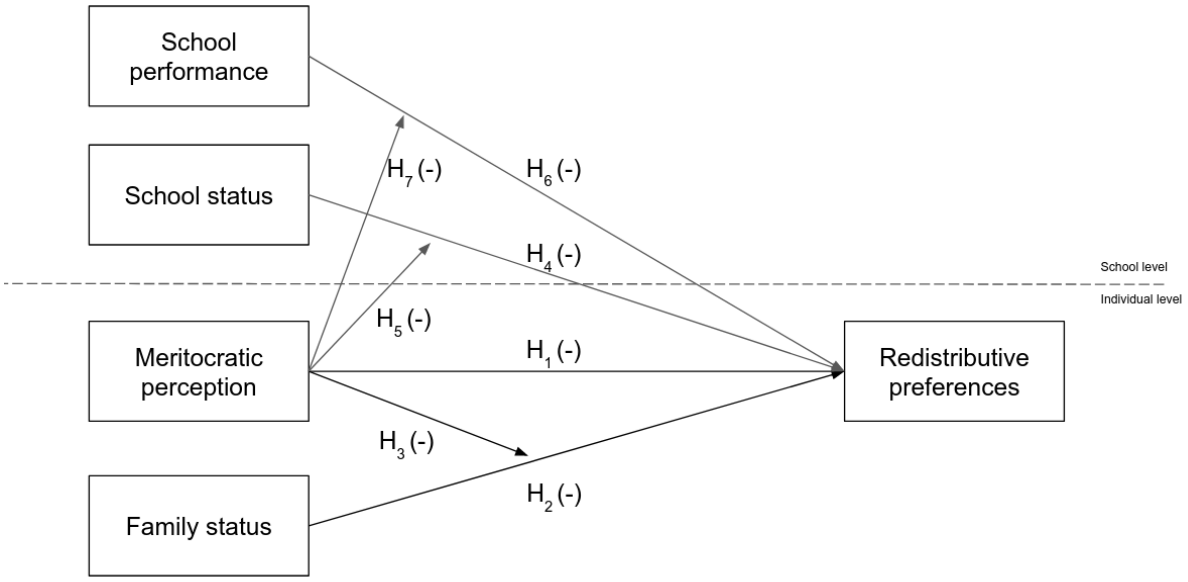
Regarding the perception of meritocracy, we will also consider two aspects of its conceptualization and measurement. In the first place, the perception of school meritocracy, associated with the fact that intelligence and effort are relevant to obtaining good grades at school (Resh, 2010). Second, the perception of meritocracy in society will be included, that is, to what extent efforts and skills are perceived to be rewarded in the country (Andersen et al., 2021; Castillo et al., 2018; Mijs, 2019).

Concerning other socioeconomic factors associated with redistributive preferences, from a rational perspective, greater demands for redistribution would be expected from those with fewer economic resources (Bernasconi, 2006; Finseraas, 2009; Franko et al., 2013; Jaeger, 2006). This would lead us to a second hypothesis that posits that **students from families and schools with lower social status would demand greater redistribution** (H_2). However, Flanagan et al. (2014) point out that in the United States, adolescents who come from more educated families have a greater understanding of wealth and poverty, while adolescents of lower social status emphasize individual explanations for the distribution of rewards, associated with the concept of meritocracy. From this point of view, our third hypothesis

indicates that **the perception of meritocracy will moderate the effect of social status on preferences for redistribution, reducing these preferences in those students with lower social status** (H_3).

About the school context, and continuing with the same sense of family status, it is expected that a **lower socioeconomic status of schools is also associated, on average, with greater preferences for redistribution** (H_4). However, this association will be **moderated by a greater perception of meritocracy** (H_5), leading low-status schools to show fewer preferences for redistribution when, on average, their students have a greater perception of meritocracy. Finally, we expect that schools with **higher average levels of academic achievement show lower redistributive preferences** (H_6), a relationship that would also be **moderated by the perception of meritocracy** (H_7), leading low-performing schools to show lower preferences for the redistribution when, on average, their students have a higher meritocratic perception.

The hypotheses are represented in the following scheme:



2. How will the crucial variables be operationalized?

State exactly how the variables specified in the hypothesis will be measured.

Our dependent variable is **redistributive preferences**, which is operationalized in two ways:

- Access to social welfare: refers to judgments about the differences in access to social welfare services according to income. The survey items are: a) “It is fair that in Chile

people with higher incomes can have better pensions than people with lower incomes"; b) "It is fair that in Chile people who can pay have a better education for their children"; and c) "It is fair that in Chile people with higher incomes can access better health care than people with lower incomes". The response categories range from 1 to 4, where 1 indicates strongly disagree, and 4 strongly agree

- Reduction of gaps: refers to preferences or support for public policies that reduce income differences between the rich and the poor. The question is: "The Government of Chile should do something to reduce the income gap between rich and poor." The response categories range from 1 to 4, where 1 indicates strongly disagree, and 4 strongly agree.

Regarding the **perception of meritocracy,** there are also two versions:

- Perception of social meritocracy: It measures the perception of rewards according to effort and intelligence separately. The survey items are: a) "In Chile, people are rewarded for their efforts"; b) "In Chile, people get what they deserve", and c) "In Chile, people are rewarded for their intelligence and skill". The response categories range from 1 to 4, where 1 indicates strongly disagree, and 4 strongly agree.
- Perception of meritocracy at school: It is proposed to measure the perception of the importance of effort and intelligence to obtain good grades. The survey items are: a) Intelligence is important to get good grades, and b) Effort is important to get good grades. The response categories range from 1 to 4, where 1 indicates strongly disagree, and 4 strongly agree.

Family **socioeconomic status** It is represented by the following variables:

- Educational level (Highest level between Respondent and Spouse/partner): Reported by parents. This variable is categorical and is encoded in the form of dummy variables (10 variables). In order not to lose cases, a category of "no information" will be included, which groups the cases of non-response from parents/guardians.
- Number of books at home: Categorical variable that identifies the number of books at home according to the student. The response categories are: 1) 0 to 10 books; 2) 11 to 25 books; 3) 26 to 100 books; 4) 101 to 200 books; and 5) 200 or more books.
- Possession of a desktop or laptop computer. The response categories are: 1) None; 2) One; 3) Two; 4) Three or more.
- Possession of Tablet or electronic readers. The response categories are: 1) None; 2) One; 3) Two; 4) Three or more.
- Possession of cell phone with internet access. The response categories are: 1) None; 2) One; 3) Two; 4) Three or more.
- Internet connection at home. Reported by parents/guardians. The response categories are 1) Yes and 2) No.

Level 2 (schools)

- Administrative dependency: Three types of school administration. The response categories are 1) Municipal; 2) Private subsidized (voucher schools) ; and 3) Non-subsidized Private. Private schools are those of higher status and represent about 8% of the immatriculation.
- Socioeconomic level according to the Ministry of Education (MINEDUC) classification. The response categories are: 1) Low; 2) Medium-Low; 3) Medium; 4) Medium-High; and 5) High
- Average in the Education Quality Measurement System test (SIMCE) of the establishment (control variable): Average of the SIMCE score of Language and Mathematics by the establishment.

3. What is the source of the data included in the analyses?

Specify the source of the obtained data. Also provide information on the context of the data source and clarify whether the data has been previously published. For simulated data, provide information on how the data was generated.

The database to be used corresponds to the First Study of Citizenship Education in Chile, carried out by the Education Quality Agency of the Ministry of Education. Eighth-grade students from 242 schools were evaluated in this study. The application date was November 9, 2017.

Three available databases are used for this study. First, there is a database of 10,213 students with identification variables (8 variables) and the civic knowledge test score (1 variable). In addition, a second database is available that contains the questions of the Citizenship Training questionnaire, which has the answers of 8,589 students (222 variables). When joining both databases, the final N is 8,589 students. Finally, a third database contains the questions of the Citizen Education questionnaire, with the answers of 6,770 parents (141 variables). When all the databases are put together, the total N of complete answers is 6,511 students and parents.

4. How will this data be obtained?

Specify how the data will be requested or accessed. Clarify if the data was already available and if the dataset has been previously explored or analyzed.

The data from the First Citizenship Education Study were requested from the Education Quality Agency of the Chilean Ministry of Education (https://www.agenciaeducacion.cl/) and are now available for our research.

There are no records of previous studies carried out with these data, except for the descriptive report of the study carried out by the Agency.

5. Are there any exclusion criteria for the data?

Specify if there were any criteria for excluding certain data sets, observations, or time points.

Cases with incomplete responses (missing cases and/or NA) will be excluded from the analyses. In addition, from the statistical package of R responsePatterns it is possible to detect repetitive response patterns, which contributes to a higher quality of research data (Gottfried, et al 2022). For example, there may be cases in which only the alternative "1" is answered or following repetitive response patterns of the type "1-2-3-4-1-2-3-4", will be excluded.

6. What are the expected statistical analyses?

Specify the statistical model that will be used to analyze the data and describe the steps prior to processing the data. Be as specific as possible and avoid ambiguity.

Before analyzing the data, the variables of interest will be recoded, and missing cases and atypical or problematic response patterns will be eliminated.

For the data analysis, first, univariate descriptive analyses will be carried out, then bivariate (correlational) analyses, and finally, multilevel models will be used to differentiate individual and contextual effects.

In the case of a high correlation between the items of preferences for access to social welfare (alpha>0.7), an average composite index will be used.

7. What are the criteria to confirm and refute hypotheses?

Specify exactly how the hypothesis will be tested. Provide specific criteria relevant to the analytical framework and model used (eg, alpha values, Bayes factor, RMSEA).

Statistically significant effects at the alpha level = 0.05 ($p \leq 0.05$)

8. Have the analyzes been validated on a subset of the data?

If yes, please specify and provide the relevant files. Please indicate whether the proposed data analyzes have been previously validated on a subset of the data or on a simulated data set. If so, please provide the data files and data syntax.

The proposed analyzes have not been validated.

9. What is known about the data that might be relevant to the hypotheses tested?

Describe any prior knowledge you have about the data set (for example, the known mean of one variable) that is relevant to your research question.

The basic descriptive statistics of the variables to be used (mean, mode and frequencies) are known.

10. Provide a brief timeline of the different steps of the previous registration.

Please provide the (expected) dates for the different steps on this pre-registration form.

-June: implementation of models and descriptions

- July-August: first draft

- September: submission to journal

Appendix

Preregistration template for the preregistration of analyzes of preexisting data

This template is based on the following article:

Mertens, G., & Krypotos, AM (2019). Preregistration of analyzes of preexisting data. *Psychologica Belgium*, *59*(1), 338-352. http://doi.org/10.5334/pb.493

Table 1. Template questions for the preregistration of analyzes of preexisting data.

Question	Description
1. What is the hypothesis that will be investigated?	*Provide a brief description of the relevant theory and formulate the hypothesis as precisely as possible.*
2. How will the crucial variables be operationalized?	*State exactly how the variables specified in the hypothesis will be measured.*
3. What is the source of the data included in the analyses?	*Specify the source of the obtained data. Also provide information about the context of the data source and clarify whether the data has been previously published. In case of simulated data, provide information on how the data was generated.*
4. How will this data be obtained?	*Specify how the data will be requested or accessed. Clarify whether the data were already available and whether the dataset has been previously explored or analyzed.*
5. Are there any exclusion criteria for the data?	*Specify whether there were any criteria for the exclusions of certain datasets, observations or time points.*

6. What are the planned statistical analyses?	*Specify the statistical model that will be used to analyze the data and describe the data pre-processing steps. Be as specific as possible and avoid ambiguity.*
7. What are the criteria for confirming and disconfirming the hypotheses?	*Specify exactly how the hypothesis will be evaluated. Give specific criteria relevant to the used analytical model and framework (e.g., alpha-values, Bayes Factor, RMSEA).*
8. Have the analyzes been validated on a subset of the data? If yes, please specify and provide the relevant files.	*Indicate whether the proposed data-analyses have previously been validated on a subset of the data or a simulated dataset. If so, provide the data files and data syntax.*
9. What is known about the data that could be relevant for the tested hypotheses?	*Please describe any prior knowledge that you have about the data set (e.g., the known mean of a variable) that is relevant for your research question.*
10. Please provide a brief timeline for the different steps in the preregistration.	*Provide the (foreseen) dates for the different steps in this preregistration form.*