Local Gender Gap Measurement in Europe: Provincial Adaptation within a Gender Mainstreaming Approach

Marco Villa[1], Daniele Toninelli[1]
[1]University of Bergamo, via dei Caniana, 2, 24127, Bergamo (Italy)

Abstract

Measures of horizontal inequality (as gender gap) are a fundamental support for planning and implementing policies. As required by the EU agenda, at all levels of policy making one should adopt a Gender Mainstreaming approach. In these terms, the Italian context can be currently analysed at the national (NUTS0) and at the regional (NUTS2) level. However, currently no tools are provided for more local analyses (i.e., at the provincial level).

The aim of this paper is to produce provincial (NUTS3) estimates for the gender gap phenomenon focusing on the Italian context. Our method relies on adapting the EU gender gap composite index (EGEI) to the provincial level. The index adaptation follows Official Statistics' standards. However, we also consider experts' critiques about the original version of such an index; this leaded us to develop additional local adaptation experiments.

Our estimates allow us on the one hand to describe gender gap heterogeneity in the Italian local context, on the other hand to study spatial relationships among different subareas. In order to develop such a spatial study, we also refer to the Official Statistics framework.

Key Words: Composite indicators, gender gap, local analysis

1. Gender Gap in Europe and the Official Statistics Role

Gender gap can be defined as the "systematic difference in the outcome of men and women on a variety of issues, ranging from economic participation and opportunity, political empowerment, and educational attainment to health and well-being" (Richardt, 2008, p. 277). As other form of horizontal inequalities, gender gap can increase the risk of conflicts in societies and leads to economic inefficiencies (Stewart, 2008; Cederman *et al*., 2011; Badgett and Hartmann, 1995). Its multi-dimensional nature requires a deep understanding of the phenomenon, that can vary also according to different contexts (Stigliz, Fitoussi and Durand, 2018).

Gender differences are intrinsically connected to citizens' welfare and should guide political decisions (Lomazzi and Crespi, 2019). With this framework, political institutions are currently aimed to mainstream the gender discussion and to increase citizens' awareness about the topic. This also holds for EU countries. Starting from the Beijing Conference (1995), the EU adopted a "Gender Mainstreaming" approach, signing political treaties that increased the Union's competence on the topic. A theoretical comprehensive definition of Gender Mainstreaming (GM) has not been produced yet (Daly, 2005), as more attention was devoted to GM approaches techniques. In this respect, some common traits can be defined among countries and institutions behaviours. In fact, GM approaches are always identified by the systematic consideration of women and men's role in communities

(Lomazzi and Crespi, 2019). Under a GM approach, gender-sensitive practices are embedded at each stage of policy making (Daly, 2005).

Local policy makers are not exempted from these mainstreamed discussions. Thus, their actions should be sustained by qualitative and quantitative measurement tools allowing the study of the phenomenon at each spatial level (from general to local contexts). Composite indexes are the main tool used for this purpose: they help in measuring the phenomenon, in detecting its evolution over time and in evaluating policies' impact. Even if composite indexes at national levels are produced and implemented in researcher analysis, alternative versions for lower administrative levels (e.g., regions and counties) are still under development (Cascella *et al.*, 2022). Unfortunately, at the time this paper was written, these local indexes did not seem to be widely used by European policy makers. This is a clear drawback, since local measure of the gender gap should also support and guide decisions and policies at such a geographical level, even to contribute to making regional/national policies more effective.

Focusing on the European area, a specific index has been developed: the European Gender Equality Index (EGEI) (EIGE, 2013). The EGEI index has been specifically built for policy making, providing yearly update about each EU country (EIGE, 2022). This index measures gender equality considering its multidimensional nature: six core dimensions are measured (work, money, knowledge, power, time and health). Additionally, two additional domains (violence and intersectional inequalities) have been included in order to provide a more complete view on the phenomenon (EIGE, 2017). Data are collected at national level and the data collection method grant an international comparison between EU countries. The final index focuses on outcome measures of gender equality, avoiding any data related to factors that produces these outcomes (i.e., time dedicated to care activities versus kindergartens availability). As will be further discussed in the methodological section, the composite index measures the equality between genders, providing values ranging between 0 (minimum equality) and 100 (maximum equality). Independent revision of the index estimation method contributed to increase the measurement quality (Permanyer, 2015; EIGE, 2017), improving the index methodology over time.

Other internationally recognized indexes have been produced; for additional information about these, see Haussman *et al.,* 2006; Branisa *et al*., 2009; Branisa *et al*., 2014; Lomazzi and Crespi, 2019; UNDP, 2022). Differences between such indexes mainly pertain the index estimation methodology and data availability. Fig. 1 presents, for the same geographical area (EU27), ranks obtained using four different indexes: EGEI, GGGI (Global Gender Gap Index[1]), GII (Gender Inequality Index[2]) and SIGI (Social Institutions and Gender Index[3]). Compared to the EGEI ranking, results can considerably vary, from one index to the other. See, for example, the cases of Cyprus, Bulgaria and Malta: for these countries the phenomenon seems to be underestimated, when an alternative to EGEI is used.

[1] See Hausmann *et al.,* 2006.
[2] The index has been produced in 2010 under the United Nations Development Programme (UNDP). The official methodology is reported on the UNDP official website (see Lomazzi and Crespi, 2019).
[3] SIGI have been produced since 2009 by the Organization for Economic Co-operation and Development (OECD). A recent revision has been made in 2019 (Ferrant *et al*., 2020).

1.1 Gender Gap at the National Level: The Italian Case

Gender Gap is a persistent phenomenon not just generally speaking (i.e., in the European Union), but even more in a specific country such as Italy. EGEI estimates currently available (referred to 2022) can help us in understanding the scale of this phenomenon. As a matter of fact, none of the EU27 member reached the score of 100 yet, that will report a full equity between genders. As reported in the annual report produced by EIGE (Fig. 2), index values range from 53, registered in Greece, to 84, observed in Sweden. At a slower pace, the EU average score has improved from 2010 to 2022 of just 5.5 points.

EGEI is not the only option available at the EU level. Indexes such as GGGI, GII and SIGI can measure the same phenomenon from slightly different perspectives (Lomazzi and Crespi, 2019). Table 1 compares, in the first column, the EGEI with the GGGI, proposed in the context of the World Economic Forum and aimed to a global comparison (Hausman *et al*., 2006). The latter adopts a broader perspective, if compared to EGEI, as it is aimed at comparing countries of different continents that subsequentially face different social challenges. Losing the European focus of EGEI, variables and dimensions slightly differ in defining and measuring the gender gap phenomenon. A practical example lies in the education dimension: in the case of GGGI, in what is defined “education and attainment” domain, the enrollment rate differences in primary education are considered as a relevant basic measure for a global comparison of gender equality. Contrarily, the EGEI focuses on tertiary education rates (part of “knowledge” domain). Differences in estimates can derive also from different data availability and data quality: in the case of the health dimension, for example, EGEI takes into account the self-perception of a good health status. This is possible because EU data producers systematically tracks this information. However, the same information is not available at a global level.

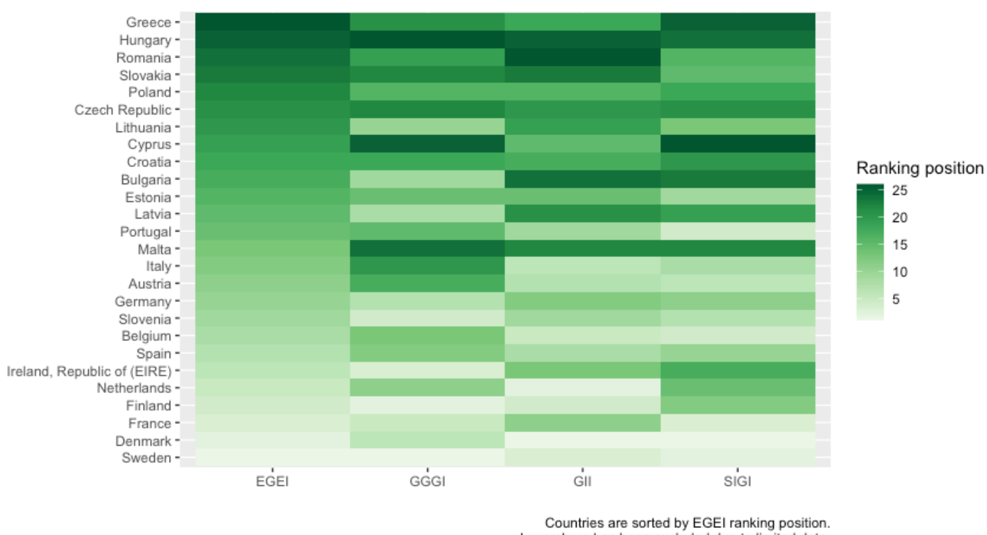


Figure 1: Ranks according to four Gender Gap indexes (EU27 members, 2018)

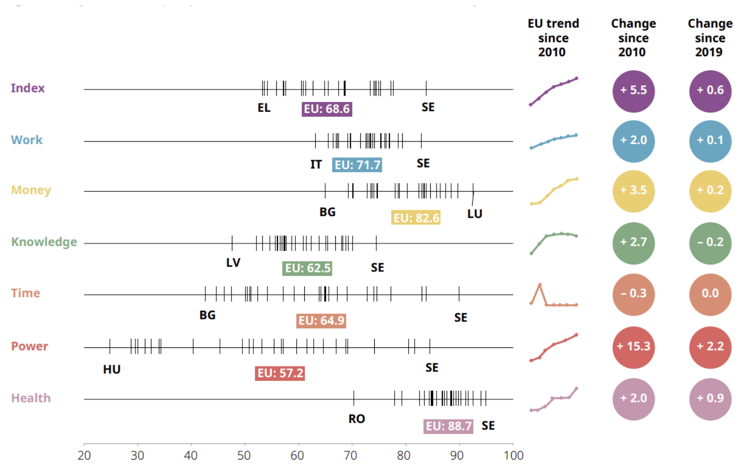


Figure 2: EGEI general level and by domain (source: EIGE, 2022)

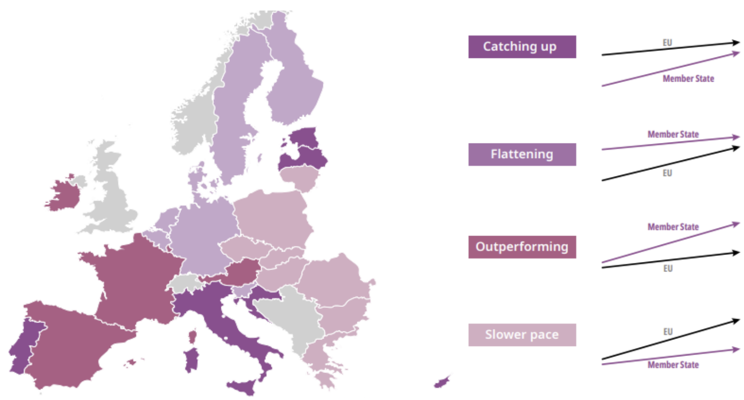


Figure 3: EGEI patterns by EU Country (2010-2022; source: EIGE, 2022

	GGGI	eRGGI	GEI (3rd edition)	R-GEI
Extended name	(Global) Gender Gap Index	extended Regional Gender Gap Index	(European) Gender Equality Index	(Regional) Gender Equality Index
Producers	World Economic Forum	Independent researchers (see ref.)	European Institute for Gender Inequality	Independent researchers (see ref.)
References	Haumann *et al.* (2006) Hausmann *et al.* (2012)	Cascella *et al.* (2022)	EIGE (2013) EIGE (2017)	Di Bella *et al.* (2020)
Level of analysis	EU and EXTRA-EU countries around the word	Italian regions	European countries	Italian regions
Measured dimensions	economic participation, education and, attainment, health and survival, political empowerment	economic participation and opportunity, use of time, educational segregation and attainment, political power and leadership, health and survival	work, money, knowledge, time, power, health	work, money, knowledge, time, power, health
Domains	4	5	6	6
Sub-domain	NA	NA	14	13
Metrics	14	31	31	25
NUTS level	Country	NUTS2	Country	NUTS2

Table 1: National and regional indexes: methods comparison.

As reported by Fig. 2, each dimension is characterized by a different trend, with the greatest increasing measured by the EU27 average for the power domain. According to Fig. 3, countries show different behaviours with respect to convergence patterns. At this regard, Italy falls into the "catching up" group, made up by countries characterised by situation improving at a faster pace than the EU average.

Fig. 4 shows the last decade measures of EGEI (Fig. 4A) and GGGI (Fig. 4B) in Italy. A clear progress is observed for this country: EGEI lifts from 53.3, in 2013, to 65, in 2022. Italian gender equality has increased by 11.7 points in the last decade, while the EU average has increased of just 5.5 points. Fig. 4A highlights the catching up pattern of Italy (black line) with respect to the EU average (red line). Note that this convergence is less pronounced, when GGGI is considered (Fig. 4B), proving that the different theoretical backgrounds (and subsequentially the different estimation methodology) are reflected in the final outcome.

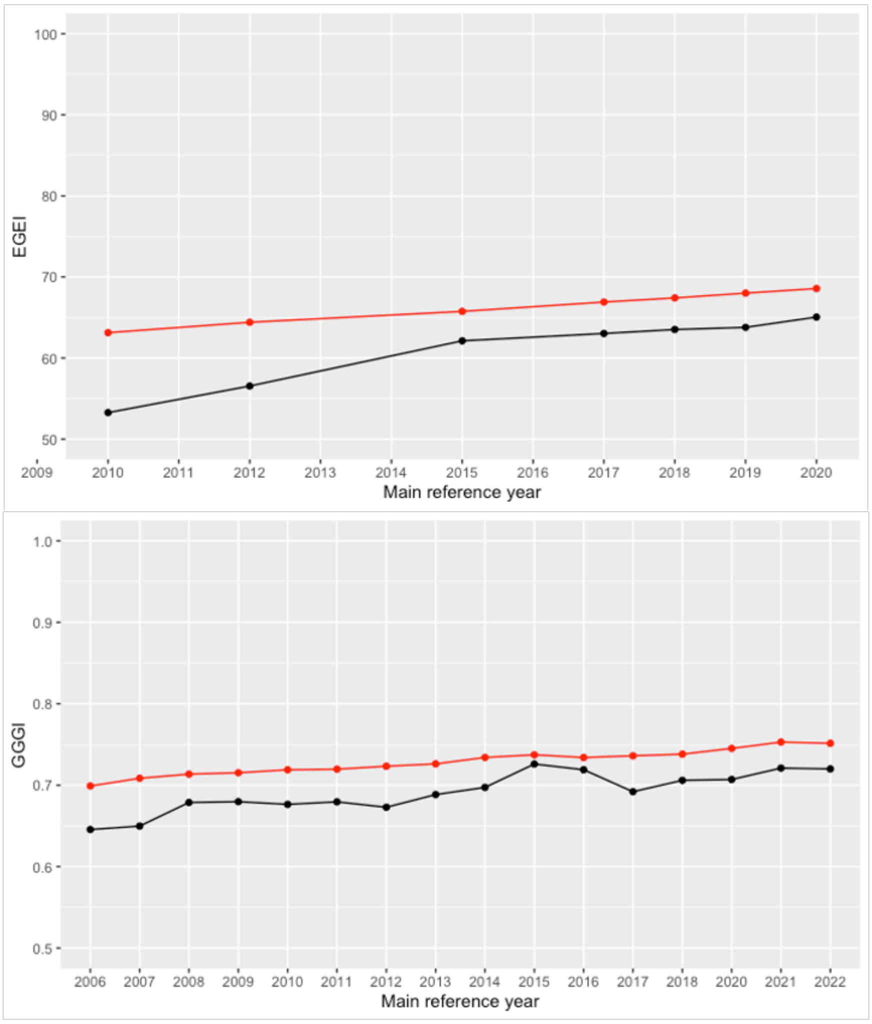


Figure 4: EGEI (A, top) and GGGI (B, bottom) estimates
(source: EGEI & GGGI dataset, 2006-2022)

Focusing just on EGEI performances (Fig. 5), Italy has closed the gap, if compared to the EU27 average, when considering the knowledge dimension (i.e., in education participation) and the power dimension (i.e., in political representation). More into detail, knowledge equality has gained 5.7 points, meanwhile power equality has increased of 31.7 points. Still, as we can see from the scale of the axis and from the general trends in EU (Fig. 1), EU levels of political equity are dramatically lower, with respect to other dimensions. In 2022, the power domain level was around 57.2, that is the lowest measure among the different considered dimensions. On the opposite side, the Italian pattern related to health is aligned to the EU27 average (90 points). Moreover, in health dimension Italy gained more (+2.7) than the EU average (+2.0). Unfortunately, both Italy and EU are far away from a gender parity, that would be reflected in an index equal to 100.

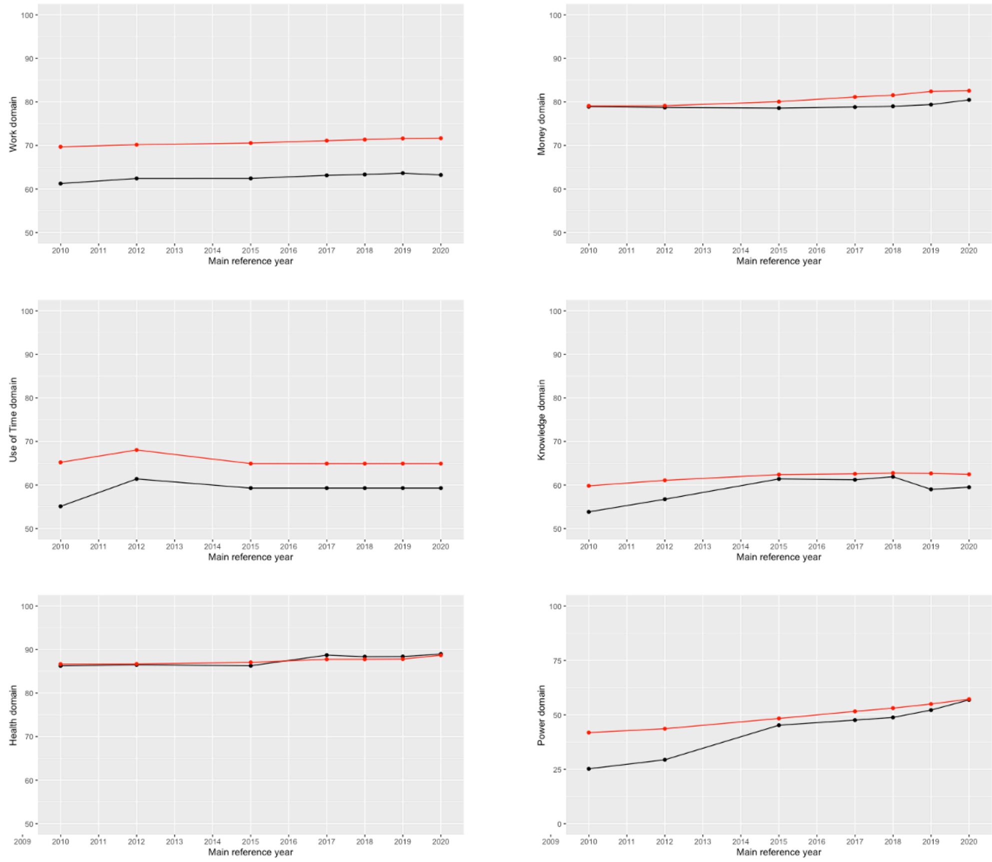


Figure 5: Italian and EU27 performances in EGEI dimensions (works, money, time, power, knowledge, health) (source: EGEI dataset, 2010-2020)

1.2 Regional and Local Adaptations

Ignoring the analysis at the local level could lead to incorrect conclusion (Duncan, 1995). Non-informative averages can be produced, especially in countries where structural differences persist (as is the case of Italy). Additionally, as we stated before, local policy making should be also locally informed to be effective. In line with GM, local tools should be provided for local policy making. These could also better inform citizens about their local context, increasing a more detailed knowledge and a deep awareness about the phenomenon and covering potential gaps and frictions between national and local issues, discussions and policies.

An index local adaptation is also encouraged by the measured phenomenon. In fact, disparities in gender gap arise when regional levels are taken into account (Di Bella *et al.*, 2020). In the European context, this can be related to gender attitudes[4]. This is

[4] Attitudes can be defined as “the beliefs regarding the appropriate roles for men and women” (Schultz Lee *et al.,* 2010; Constantin and Voicu, 2015).

confirmed also by regional studies, that reported a larger regional variability in European gender attitudes with respect to national variability (Cascella *et al.*, 2022).

At the current stage, two local index adaptations that are specifically suited for Italy has been published (Di Bella *et al*., 2020; Cascella *et al*., 2022). Both the adaptations produce an index at the NUTS2 level: this allows to study Gender Gap regionally. In both works, authors highlight the main difficulties of locally adapting national indexes. The main limit is linked to available data and their representativeness. This is a structural limit of Official Statistics Framework, that imposes a specific NUTS level representativeness in each data production process. Of course, new indicators can be added, when local alternatives of original variables are not available (Cascella *et al.,* 2022), but this could cause a biased estimate of the index itself.

What about more local levels? How to produce a more detailed (but still reliable) index allowing to obtain a more local evaluation of the phenomenon? Filling this objective is the main scope of this paper. In this work we aim at producing a Gender Gap index specifically suited for the local Italian context at the NUTS3 level. From now on, we will refer to this new index as the Provincial Gender Gap Index (PGEI). However, this work also aims at encouraging other researchers to regionally adapt Gender Gap indexes for their local contexts. The decision of adapting a national index to the NUTS3 level is not randomly taken. Besides of local administrations, NUTS3 level determines, in Italy, one of the most relevant geographical level of local policy making. Italian Central government representatives ("prefetture") have NUTS3 competences in fields as education and security. Moreover, autonomous NUTS3 regions exists and are endowed with the same competences of ordinary NUTS2 policy makers. Additionally, Italian larger urban areas centrally manage (at NUTS3 level) infrastructure investments and territorial planning.

2. Data and Methodology

This section introduces our dataset and presents the methodology used in order to locally adapt the gender gap index. Our method is based on the replication of the original methodology proposed in EIGE (2013). Nevertheless, we implemented some changes following a data driven approach: we highlighted and tried to validate the reasons behind each of our choices, within the adaptation procedure.

This section continues with a brief description of the studied data (section 2.1). Then, we introduce the algorithm used for estimating basic metric (section 2.2). Section 2.3 is focused on domains and sub-domains validation and their internal consistency. Section 2.4. evaluates potential sources of uncertainty in the index estimation. In this same section, the best performing index will be compared to its national counterpart.

A short note: with the term "indicators" or "variables" we refer to the original variables later used in computing the basic measures of gender gaps. These variables can be expressed in different scales or units. Instead, with the term "metrics" we refer to any normalized version of a variables / indicators.

2.1 Data and main technical change

The identification of the main variables to compute the overall gender gap index is based on the theoretical framework developed by the European Institute for Gender Equality (EIGE). The list of final indicators / metrics and the considered source datasets are shown in Tab. 2. For each dataset, we applied data imputation methods, in order to obtain a complete dataset including 107 provincial observations for each gender (M/F) and for each year (2018 and 2019)[5]. Data were collected between January and May 2022.

The final dataset includes 14 different metrics, divided into 5 domains: work, money, knowledge, health and power. Indicators where originally divided into subdomains: participation and segregation (work), financial resources and economic situation (money), participation and segregation (knowledge), political and economic (power). In health domain, only one variable was considered, defining the health status of the population through life expectancy. Due to data unavailability, it was not possible to retrieve neither the original variables nor some proxy for the domain of "use of time".

Data were collected by five different institutions belonging to the Italian statistical system (SISTAN), the network of public and private institutions that produces official statistics information. Data collection lasted from February 2022 to June 2022. All observations, if necessary, have been recoded into annual periodicity.

Some choices were made to improve the quality of the final index. In particular, in the following we list the main ones.

- Metrics *power_1.2* and *power_1.3* only consider municipalities including less than 15,000 inhabitants that are managed under ordinary administration. The final variables have been created considering the administered population by each political figure. The choice of limiting the analysis to this cluster of municipalities arose following a preliminary statistical analysis. When municipalities with less than 15,000 are considered, we observe a higher linear correlation between local female political representativeness and the number of inhabitants. Furthermore, weighting political representation by municipality size allows to take into consideration a gender element of the local political culture.
- Metric *work_1.2* (complementary to the non-participation rate in work) was included to further strengthen the work domain of participation.
- Metric *work_2.1* shows the future job vacancies divided by gender. preferences as declared by the employer. Preferences are divided into male, female and indifferent gender. The indifference class was equally reassigned to female and male preferences.
- Metric *know_2.2* (students with unsatisfactory literacy levels) has been included as a proxy for segregation in higher education programs. This choice was determined considering the lack of detailed data by gender and province of enrolled students. At the same time, the choice of this proxy has found confirmation in the literature of the barriers generated by low numerical skills in further accessing STEM paths (Bakker *et al.*, 2018; Borgonovi *et al.*, 2021).

Index estimation also required choices in terms of data imputation. More specifically, we implemented the following strategy.

- *Provincial administrator's registry* (obtained from the Italian *Ministry of the Interior*)
  At the time of data imputation, an issue was linked to the autonomous regions of Friuli-Venezia-Giulia, Sicilia, Sardegna and Valle d'Aosta. Friuli-Venezia-Giulia, Sicilia, Sardegna reformed their provincial administrative, while Valle d'Aosta region does not have a provincial level of analysis.

[5] More information will be provided in the following paragraphs.

In the case of the Friuli-Venezia-Giulia, the administrative tasks have been redistributed among municipalities and regional bodies. For this purpose, provincial indicators were replaced by the average of regional and municipal level indicators.

Sicilia and Sardegna (excluding the Cagliari province) were still missing their provincial bodies. For this reason, the provincial indicators were computed on the potential citizens of provincial administrators that could join those bodies, taking into consideration the political weight of the municipal administrators.

For Valle d'Aosta, the provincial index was calculated on the regional indicators.

- *BES indicators (ISTAT), CAMCOM Marche (InfoCamere), Private Worker's Observatory (INPS) and Excelsior (Unioncamere)*

  In these datasets, the provinces of South Sardegna, Monza and Brianza and Fermo were missing due to delays in updating the official province classification. Actually, when missing, the data provided included the results of these "new" provinces in the reference provinces. For example, the province of Milan incorporates the data regarding the Monza and Brianza province; Cagliari incorporates data for South Sardegna; Ascoli Piceno data includes figures for Fermo. In such cases, the value of the origin province was applied also to the new province.

  A different situation applies to Foggia and Bari: both provinces transferred part of their territory to create a new one (Barletta-Andria-Trani). In this case, since both original provinces are still active, it is possible to consider the reference populations to redistribute the original provinces' values into the new ones.

Lastly, it should be considered that each indicator, when necessary, has been linked to a specific reference population. This happens, for example, for the indicator *work_2.1*, that has been created considering a specific workforce subpopulation (15-74 years).

Most of the indicators are normalized considering the maximum achievement obtained by all provinces in a specific measure. This way, the index actually takes into account the distance between each value and the gold standard that have been measured in Italy until that moment. In most cases the achievement is based on the same original value. In other cases, the achievement variable has been replaced by more suitable variables. This happens, for example, in the following cases:

- *work_2.1*, where we used the total future job vacancies in the considered sector;
- *power_2.2*, where we considered the share of active companies with more than 50 employees on the total number of active enterprises.

Thanks to this approach, it was possible to partially consider the entrepreneurial demography of each territory.

Domain	Sub Domain	Adapted final metric (reference population if needed)	Recoding	Data source
WORK	Participation	Employment rate (% 20-64 population)	*work_1.1*	ISTAT - BES
		Nonparticipation rate (% 15-74 population)	*work_1.2*	ISTAT - BES
	Segregation and Quality	Share of gender-defined future job vacancies in education, health, social assistance, cultural & sport sectors and other services (% 15-74 population)	*work_2.1*	Excelsior dataset
MONEY	Financial Resources	Average annual work income of private sector employees	*money_1.1*	ISTAT - BES
		Average annual income from social security welfare	*money_1.2*	ISTAT - BES
	Economic Situation	S80/S20 declared income quintile shares	*money_2.1*	Ministry of Economy and Finance of Italy
KNOWLEDGE	Attainment and Participation	Rate of tertiary educated population (25-49) (% 25-49 population)	*know_1.1*	ISTAT - CensPop
	Segregation	Ratio of students with unsatisfactory achievements in numeracy levels (% last year high school students)	*know_2.2*	ISTAT - BES
HEALTH	Status	Life expectancy in absolute year at birth	*health_1.1*	ISTAT - BES
POWER	Political	Share of administrator (president + others) in provincial councils (% 20+ population)	*power_1.1*	Ministry of the Interior
		Share of mayors, weighted for administered population (% 20+ population)	*power_1.2*	Ministry of the Interior
		Share of municipality administrators (mayor excluded), weighted for administered population (% 20+ population)	*power_1.3*	Ministry of the Interior
	Economic	Workers classified as managers in the private sector (% workers in private sector)	*power_2.1*	Private Worker's Observatory
		Share of active companies (%15-74 population)	*power_2.2*	Infocamere

Table 2: Adapted variables from EGEI methodology.

2.2 Basic metrics creation

Each basic indicator requires a transformation process to obtain a provincial measure of gender gap. This method is introduced by formulas 1 to 4. As preliminary step, we identify the polarity of the index for each indicator. In this regard, an indicator is defined as positive if the increase in its value corresponds to an improvement in the phenomenon under investigation. In the case of indicators with negative polarity, in order to avoid compensation problems during the aggregation phase, the index was rather reversed. In case of missing values, we implement the mentioned imputation strategy.

Most indicators need to be expressed in relative terms to compare populations with different sizes and structures (equation 1). This recoding is not necessary for indicators expressed in per-capita monetary values, such as the average income indicators. However, some indicators are already expressed in relative terms (i.e., employment rate and non-participation in work), consequently they did not need to be transformed.

$$\tilde{x}_{i,t}^{k} = \frac{x_{i,t}^{k}}{reference\ population_{i,t}^{k}} \quad (1)$$

Where:
$\tilde{x}$ = variables expressed in relative term;
k = gender (woman, man, average of the two);
t = time (2018, 2019);
i = province;
x = original variable.

As shown by equation 2, the normalization procedure generates a gender-neutral measures that considers the general level of achievement for Italian provinces (equation 3). In this step, values are:

- compared to the total average value of the two genders;
- benchmarked to the full equity situations (corresponding to 1);
- considered in absolute terms, so that none of the genders would dominate the analysis.

$$\gamma_{x_{i,t}} = \left| \frac{\tilde{x}_{i,t}^{w}}{\tilde{x}_{i,t}^{ave}} - 1 \right| \qquad (2)$$

Aiming at producing a measure of the total achievement, each metric is corrected by the maximum level reached among all provinces ($\boldsymbol{\alpha}$); this value is then compared to the maximum level reached during the two years among all the territories (equation 3). This adjustment was included starting from the 2017 version of the index, in order to reduce the weight of the achievement component on the variability of the index (Permanyer, 2015). Finally, note that this correction is not applied when the basic indicator represents quotas set by law, as the maximization of quotas would be conceptually wrong. This specific situation occurs in the indicators referring to the power dimension, where the gender representatives are imposed by the legislator.

$$\alpha_{x_{i,t}} = \left(\frac{\tilde{x}_{i,t}^{ave}}{max\{\tilde{x}_{i,2018}^{ave}, \tilde{x}_{i,2019}^{ave}\}} \right)^{\frac{1}{2}} \qquad (3)$$

For a better interpretation of results, we compute the inverse value of the calculated ratio (γ). The result is an index with ranging from 1 to 100, where the gender parity corresponds to 100 and an absolute gender difference is set at 1 (equation 4).

$$\Gamma_{x_{i,t}} = 1 + \left[\alpha_{x_{i,t}} * (1 - \gamma_{x_{i,t}})\right] * 99 \qquad (4)$$

The final metric will subsequently be aggregated to estimate the individual domains. Aggregation by domains will take place through a first aggregation by subdomains. Lastly, domains will be aggregate to obtain the Provincial Gender Equality Index (PGEI).

2.3 Aggregation and consistency

To verify the correctness of each sub-domain definition, the EIGE methodology suggests starting using the results of the Principal Component Analysis (PCA). At this stage, PCA can be implemented without a further variable standardization, since data have already been normalized into a 1 to 100 range. However, a proper standardization process requires that both average and standard deviation are the same for all metrics. Thus, we applied a further standardization, in order to purge the results of any unwanted variance-driven effect. To improve the analysis, 2018 and 2019 observations were jointly considered for each province.

We also checked the internal consistency of the index domains computing the Cronbach's alpha on the entire set of metrics and on individual domains. Furthermore, to check the effects of each metric on the internal consistency, it is suggested to compare the alpha value obtained on the entire dataset to the alpha value calculated in the event of exclusion of the metric (Cascella *et al.*, 2022), Finally, we highlight that this evaluation is not affected by the sources of uncertainty that will be introduced in the next section (2.4).

2.4 Estimating the final index

The Index estimation involved an assessment of each possible sources of uncertainty. The main uncertainty sources include the following choices:

- the basic indicators aggregation method (arithmetic, harmonic or geometric mean);
- the domain aggregation method (arithmetic, harmonic or geometric mean);
- the weights used in the aggregation of basic metrics (no weight or weights deriving from the PCA);
- the weights used in the domain aggregation (no weight or redistributed AHP[6] weights);
- division into subdomains (following the theoretical framework or following the statistical results).

This combination of choices produced 72 final alternative indicators. In order to identify the best one, we minimize the Euclidean distance between the median of all indicators and each estimate in all provinces (equation 5).

$$d_j = \min_j \sqrt{\sum_{i=1}^{107} (I_{i,j} - I_{median_i})} \quad (5)$$

where:

j = estimated index (j = 1, …, 72);
i = provinces (i = 1, …, 107).

2.5 Spatial assessment of index versions

Describing the spatial behaviour of our index is fundamental to properly allow the usage of this information in econometrics models. Its relevance arises from the nature of our data, that are inherently defined by their geographical nature of "local" information. Again, official statistics propose a specific, commonly used methodology in order to spatially describe our data (see also Audric *et al.*, 2018).

The Moran I index allows to assess the global correlation level observed among our geographical data. We computed this index for each of the 72-gender gap index versions. This methodology requires a definition of the weighted matrix that a-priori better describes the spatial relationships. Following ISTAT (2019), we adopted a queen matrix with line standardization. Network connections can be observed looking at Fig. 6. The standardization line applied to the queen matrix increases the results interpretability: each weight represents the fraction of spatial influence on observation i ascribed to the relationship between observation i and j (ij). The sum of weights should equal 1, i.e.:

$\sum_{j=1}^{n} w_{ij} = 1$ (Belleon *et al.*, 2018)

[6] As expressed in the EGEI methodology, the "Analytic Hierarchy Process (AHP)" relies on experts' evaluation to derive the importance of each measure (EIGE, 2013).

As suggested by Salima (2018), we test two different calculation approaches, based on Monte Carlo Simulations and on randomic analytical results. The analysis has been developed using the SPDEP R package (2023-02 version), following the suggestion of Bivard (2022)[7].

Figure 6: Network connections between provinces detected using *Queen* contiguity principles

2.6 Local cluster analysis

In order to evaluate the local heterogeneity in the spatial relationships between areas, we developed a local analysis basing on the class of Local Indicators of Spatial Association (LISA) (Loonis, 2018; ISTAT, 2019). In particular, we computed the Local version of the Moran Index, which guarantees the proportionality to the aforementioned global version of the index (Anselin, 1995). To conduct this analysis, we limited our dataset to the 2018 version of the computed Provincial Gender Gap Index (PGEI).

The methodology suggests preferring a computational approach over an analytical approach (Loonis, 2018). This can be done thanks to conditional permutations of the included information. In practice, PGEI values for the considered *i*-th province will be held constant, while values of neighbours will be randomly permuted with other provinces values. Reiterating the same provincial information in multiple tests generates risk inflation. We controlled for this issue introducing the Bonferroni correction, which adjusts the alpha value considering the number of statistical tests to be executed (equal to the number of provinces and increased by 1). This correction is sometimes considered too restrictive (Loonis, 2018). Thanks to the function spdep::p.adjustSP we adjusted the *p*-value only for the relevant provinces (included in

[7] We set the main function parameters as follows: number of permutations = 999, hypothesis test = greater hypothesis, seed = 111.

the neighbours)[8]. Also in this case, all the computation has been made using R and the SPDEP package[9].

In order to detect potential clusters of provinces, we categorize our results considering the correlation direction (positive or negative) and the PGEI value observed for each province (above or below the national mean value). Local clusters will be defined as high / low levels that are positively correlated with their neighbours. Instead, we will characterize local outliers as provinces negatively correlated to their neighbours. As suggested by Anselin (2016)[10], we focus on different p-values levels to identify the nucleus of a cluster and its surrounding neighbours. As proposed by him, and only in the case of positive correlation, a non-significant correlation of a neighbour can still be useful to definie the surrounding area of a significant nucleus. For this reason, in our results we will present different levels of alpha that have been considered.

3. Results

In this section we present the results of our study. An exploratory data analysis on the basic metrics is conducted in section 3.1. Uncertainty level results are presented in section 3.2. Section 3.3 compares our index (PGEI) performance to the original index (EGEI) and to the NUTS2 adaptation (RGEI). In section 3.4, we further check the spatial patterns of all indexes we consider. Finally, we provide an overall description of gender gap at NUTS3 level has been conducted (section 3.5), including the definition of spatial cluster at the Italian provincial relevel (section 3.6).

3.1 Preliminary analysis

Table 3 shows the summary statistics of the final gender gap metrics, jointly considering 2018 and 2019 data. It should be kept in mind that a higher value of the metric corresponds to a lower gender difference. Therefore, a value of 100 corresponds to full parity between genders, while a value of 0 corresponds to a full disparity between genders, regardless of which is the dominant gender.

As highlighted by the average and median values, each metric presents a different situation. The lowest median value is observed in power dimension, whereas it is in health dimension that the highest median value can be found. Besides the health domain, standard deviations suggest a high variability in all other dimensions. Skewness and kurtosis suggest that all metrics can be considered normally distributed, with the exception of *power_2.1*. This metric shows a longer right tail (skewness = 2.18) and a greater presence of observations in the tails (kurtosis = 9.7), highlighting an uncommon level of inequity in the private sector management.

Correlations between metrics are generally highly significant and positive (Tab. 4). This result is in line with the results obtained by EGEI and suggests a high consistency and a correct composition of the final measure. The highest correlations do not always occur within domains. This is the case of the metrics referring to high school population with a sufficient level of numerical skills (*know_2.2*). This metric is highly correlated with metrics of the work domain. High inter-domain correlations also occur among metrics belonging to the power domain and among metrics belonging to other domains (such as the knowledge domain and the money domain). These results, although limited

[8] The correction factor m of Bonferroni correction will be different by cluster, where we will consider the number of neighborhood provinces for each cluster. I.e. *m* parameter in spdep::p.adjustSP, for Milan cluster (with 7 neighbors), will be set equal to 7 (Anselin, 1995).
[9] We set the main function parameters as follows: number of permutations = 999, hypothesis test = two-sided hypothesis, seed = 111.
[10] The relevant part is introduced in Anselin lectures, minutes 20:00 (Anselin, 2016).

to a linear interpretation of the phenomenon and without providing causality information, support the need for a further multidimensional analysis of the phenomenon, as they show a strong statistical inter-relationship between metrics.

Metric	**N. of observations**	**Mean**	**SD**	**Min**	**Q25**	**Median**	**q75**	**Max**	**Skew**	**Kur**
health_1.1	214	96,56	0,61	94,91	96,10	96,60	97,06	97,78	-0,27	2,47
know_1.1	214	65,35	6,32	44,11	61,21	65,11	68,62	87,32	0,40	4,59
know_2.2	214	79,31	10,39	51,79	73,57	81,69	87,01	97,75	-0,59	2,42
money_1.1	214	63,93	6,90	49,75	58,37	63,70	69,09	83,20	0,21	2,50
money_1.2	214	75,01	2,89	66,95	73,04	74,79	77,06	83,05	0,18	2,76
money_2.1	214	72,57	17,10	24,19	62,73	78,05	86,47	95,19	-0,89	2,93
power_1.1	214	42,58	22,59	1,02	22,96	44,74	59,22	95,18	-0,12	2,18
power_1.2	214	27,71	17,11	1,00	15,56	27,00	36,49	78,94	0,62	3,41
power_1.3	214	65,12	8,52	43,41	60,02	65,92	70,82	91,71	0,00	3,31
power_2.1	214	30,07	9,39	16,57	24,22	28,33	32,46	73,73	2,18	9,67
power_2.2	214	30,06	5,08	19,23	26,50	30,43	33,95	43,19	-0,11	2,32
work_1.1	214	75,07	13,03	44,77	64,45	80,82	84,93	92,90	-0,77	2,31
work_1.2	214	86,58	9,60	60,70	79,95	91,34	93,80	99,43	-0,96	2,68
work_2.1	214	54,52	10,68	34,27	46,35	53,57	60,97	86,12	0,63	3,26

Table 3: Basic metrics summary statistics (2018 and 2019)

	health_1.1	know_1.1	know_2.2	money_1.1	money_1.2	money_2.1	power_1.1	power_1.2	power_1.3	power_2.1	power_2.2	work_1.1	work_1.2	work_2.1
health_1.1	1.000													
know_1.1	0.431***	1.000												
know_2.2	0.485***	0.416***	1.000											
money_1.1	0.446***	0.445***	0.688***	1.000										
money_1.2	0.204**	0.513***	0.333***	0.581***	1.000									
money_2.1	0.361***	0.233***	0.661***	0.704***	0.599***	1.000								
power_1.1	0.086	-0.127	0.090	0.235***	0.148*	0.180**	1.000							
power_1.2	0.393***	0.265***	0.407***	0.438***	0.309***	0.385***	0.246***	1.000						
power_1.3	0.206**	0.107	0.241***	0.381***	0.242***	0.348***	0.351***	0.284***	1.000					
power_2.1	0.274***	0.557***	0.318***	0.766***	0.540***	0.331***	0.155*	0.306***	0.324***	1.000				
power_2.2	0.268***	0.349***	0.644***	0.526***	0.276***	0.419***	0.059	0.288***	0.248***	0.250***	1.000			
work_1.1	0.483***	0.415***	0.782***	0.798***	0.603***	0.822***	0.254***	0.503***	0.346***	0.437***	0.526***	1.000		
work_1.2	0.481***	0.362***	0.827***	0.779***	0.512***	0.804***	0.194**	0.491***	0.331***	0.399***	0.540***	0.971***	1.000	
work_2.1	0.370***	0.348***	0.350***	0.390***	0.282***	0.288***	0.145*	0.286***	0.172*	0.287***	0.153*	0.502***	0.490***	1.000

Note: *** = pvalue < 0.001 , ** = pvalue <0.01 , * = pvalue < 0.05

Table 4: Pearson's correlation matrix of final metrics

3.2 Internal consistency of domains and subdomains

As is highlighted by the single loading factors of each principal component, subdomains are correctly defined only in work domain, correctly dividing into two sub-domains the three available variables[11]. Looking at the normalized loading factors of knowledge domain[12], results suggest that a subdivision between subdomains is not necessary. Power domain also presents an anomalous situation, where the scree plot suggests creating a third subdomain[13]. About this domain, even limiting the analysis to the first two principal components, with a cumulative explained variance around 80%, the division proposed by the factor loadings would not follow the expected division proposed by the framework. Focusing on money domain[14], PCA results suggest a new classification. The data-driven suggestion includes in a single domain the metric referring to the income from private sector employees (*money_1.1*) and the metric

[11] See Annex 1.
[12] | 0.71| for metric in both components. See Annex 2.
[13] See Annex 4.
[14] See Annex 3.

referring to the upper / lower quintile share (*money_2.1*). At the same time, this solution leaves in a separate extra domain the income derived from social security welfare (*money_1.2*).

However, as a good practice, statistical considerations cannot ignore a practical / actual meaning. As regards power domain, given the weakness of the directions reported by the biplot, we should follow the theoretical framework and keep two power sub-domains (economic and political). Work domain presents a double solution too. Considering these results, it was decided to insert a new source of uncertainty in the index creation. This allows to evaluate whether the choice to make the theoretical framework prevail allows to provide with better results than the choice based on aggregation suggested by PCA. In this respect, only for the knowledge domain the PCA results were not assessed, and two sub-domains (segregation and participation) were taken for granted.

Considering the entire set of metrics, internal consistency of the index is higher (0.84), if compared to the one measured for each domain. Alpha values for single domains vary between 0.53 (power domain) and 0.84 (work domain)[15]. Alpha values, when an indicator is excluded, do not seem to change considerably; they rather remain always above the golden standard. *Power_1.1* (share of political figures in the provincial bodies) is the metric that, once excluded, mostly improves the internal consistency of the index. On the contrary, the lowest internal consistency is obtained excluding *work_1.2* (inverse of the rate of non-participation in work), with an alpha value equal to 0.77 when the lower bound of the confidence interval is considered. Work domain has the highest alpha (0.843), closest to the full set measurement. Further analyzing this domain, there is an improvement of internal consistency when *work_2.1* (gender quotas in the personal services sector) is excluded. However, there is not a clear suggestion to exclude this indicator since, once excluded, the internal consistency of the composite index is basically unchanged[16].

3.3 Index estimation

Compared to the original methodology, PGEI is different both in terms of weights and aggregation method with respect to the original EGEI. The decision rules of equation 5 lead to an indicator built following the next criteria.

- **weights**: PCA for metric aggregation, equal weights for sub-domain aggregation, equal weights for domain aggregation.
- **aggregation**: weighted arithmetic means for each level of aggregation.
- **reference for aggregation**: theoretical framework.

To check the final quality of the new index, EGEI methodology suggests comparing the ranking defined by the best scenario to those defined using other scenarios (EIGE, 2013). A simple plot can identify the variability of gold standard index, with respect to the discarded version. As shown by Fig. 7, there is a strong concentration of observations around the zero-change value. Studying the cumulative frequencies, about 13% of the observations did not show a shift in the ranking. This percentage increases by 42%, if we consider variations between -2 and + 2. By extending the analysis to variations between -4 and + 4, the included observations are more than 60%. Compared to the original index (EGEI), a lower robustness is measured by PGEI, where 85% of the observations are included between -2 and + 2 (EIGE, 2013).

[15] See table 5.
[16] From 0.843 (when included) to 0.835 (when excluded). See table 5, 6.A, 6.B and 6.C.

DOMAIN	EXCLUDED	Cronbach's alpha	C.I. Lower Limit	C.I. Upper Limit
	None	0,843	0,813	0,869
HEALTH	*health_1.1*	0,846	0,816	0,872
KNOWLEDGE	*know_1.1*	0,839	0,808	0,865
	know_2.2	0,820	0,786	0,850
MONEY	*money_1.1*	0,822	0,788	0,851
	money_1.2	0,839	0,808	0,866
	money_2.1	0,816	0,780	0,847
POWER	*power_1.1*	0,877	0,856	0,895
	power_1.2	0,833	0,800	0,861
	power_1.3	0,835	0,804	0,863
	power_2.1	0,831	0,800	0,859
	power_2.2	0,837	0,805	0,864
WORK	*work_1.1*	0,803	0,765	0,836
	work_1.2	0,813	0,777	0,844
	work_2.1	0,835	0,803	0,862

Table 5: Cronbach's alpha for index

DOMAIN	EXCLUDED	Cronbach's alpha	C.I. Lower Limit	C.I. Upper Limit
MONEY	None	0,624	0,583	0,663
	money_1.1	0,329	0,282	0,379
	money_1.2	0,656	0,602	0,706
	money_2.1	0,585	0,513	0,654

DOMAIN	EXCLUDED	Cronbach's alpha	C.I. Lower Limit	C.I. Upper Limit
POWER	None	0,532	0,445	0,616
	power_2.2	0,524	0,430	0,615
	power_2.1	0,471	0,377	0,567
	power_1.1	0,542	0,457	0,624
	power_1.3	0,433	0,332	0,540
	power_1.2	0,412	0,315	0,515

DOMAIN	EXCLUDED	Cronbach's alpha	C.I. Lower Limit	C.I. Upper Limit
WORK	None	0,843	0,802	0,878
	work_2.1	0,962	0,956	0,968
	work_1.1	0,655	0,559	0,740
	work_1.2	0,659	0,566	0,742

Table 6: Cronbach's alpha for Money domain (6.A), Power domain (6.B) and Work domain (6.C)

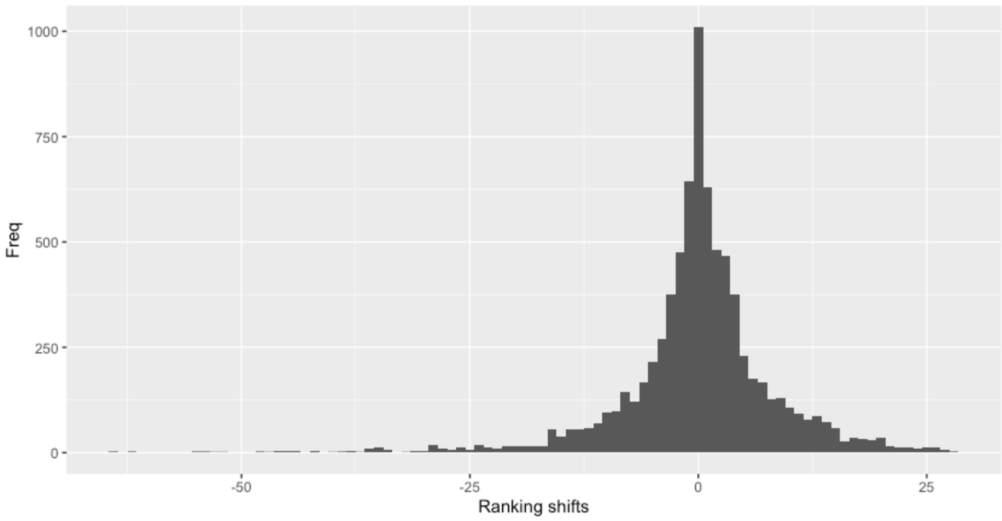


Figure 7: Histogram of ranking shifts, measured on 72 scenarios for each province vs gold standard scenario.

Finally, in order to further assess the quality of results, we studied the linear correlation between the final index and the considered domains and subdomains. As can be seen in Annex 5, the strongest and most statistically significant positive correlations usually occur between domains and the index[17]. As EIGE suggests, these correlation values show that an optimal analysis can be conducted even using a limited set of sub domains in further gender gap studies (EIGE, 2013). Furthermore, this strong correlation is a signal of how each subdomain actually contributes to the measurement of the phenomenon (EIGE, 2013).

In almost all the cases, and as a direct effect of PCA-based procedures[18], subdomains measure the highest correlation with their own reference domain. Only in the case of *power_2* subdomain (economic power), the highest correlation is measured with the domain related to knowledge[19]. This subdomain is also characterized by the highest correlation with the subdomain related to financial resources. This result suggests that omitted variable bias could affect this result, as there is not a clear logic for this strong relationship. In fact, the positive strong correlation between gender gap for unsatisfactory numeracy skills in high school (*know_2.2*) and gender gap in entrepreneurship (*power_2.2*) is not straightforward. Hypothetically, it could happen that numeracy skills could discourage the later enrollment to learning path focused on management. This would limit both the competencies and the networking activities that are helpful for future entrepreneur.

A further check can be conducted comparing EGEI results to PGEI results, when the newly created index is measured at the national level. If compared to PGEI, EGEI index shows a better overall situation both in 2019 (70.56 vs 63.78) and in 2018 (69.74 vs 63.51). Relatively higher values for EGEI domains can also be found in health, knowledge and work dimensions. Conversely, PGEI index shows a worse situation, with greater gender gap in power and money domains. Differences between indexes and domains are generally below 10 points, except for the knowledge and the power domains. Results obtained from the aggregate measure, with differences between 6 and

---

[17] I.e., correlation between PGEI and money domain is 0.889, while individual subdomains correlation with PGEI is respectively 0.868 (*money_1*) and 0.818 (*money_2*).

[18] See section 3.2.

[19] Correlation between subdomain *power_1* and domain power are 0.616, while the one between *power_1* and knowledge domain is 0.708. See Annex 5.

7 points, could be a good sign of the successful estimation of the index at the provincial level. However, it should be noted that these discrepancies are higher than the two-point one observed between EGEI and the regional adaptation R-GEI.

3.4 Spatial assessment of index versions

Fig. 8 shows some measure of the performance of the estimated indexes. Global spatial correlation is positive and statistically significant ($p < 0.001$) for each of the 72 versions. Global Moran's index results are consistent over index versions and calculation methodologies. As showed by Fig. 8, the randomic approach produced a left-skewed distribution of correlation indexes. Moran's I values are always positive and included between 0.65 and 0.8. Similarly, Monte Carlo simulations produced not acceptable results under the null hypothesis of non-spatial correlation between PGEIs. In other word and globally speaking, nearby provinces tend to share the same level of gender equality, and this is true across all index versions (not just the gold standard version). Fig. 8 show the Moran's I plot for our data, that is a simple tool able to show the association between provincial results. For each observation, the average information for its neighboring observation[20] are reported. Aligned with the positive value of the global correlation index, most of the NUTS3 information lays on high-high and low-low correlation areas.

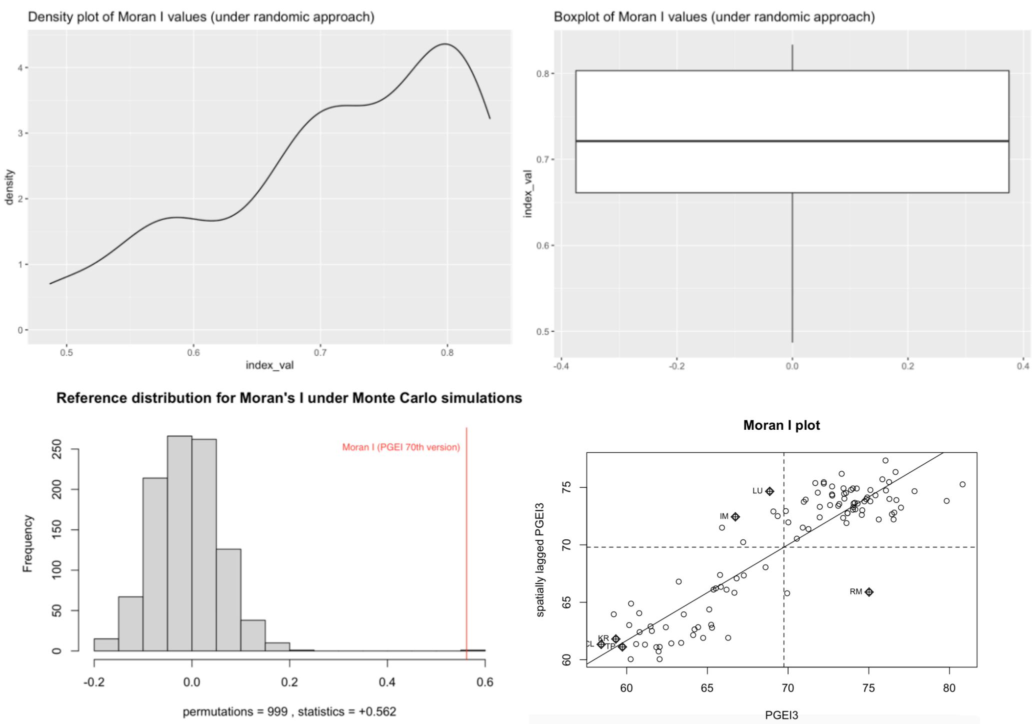


Figure 8: Summary performance report of Global Moran I

3.5 Gender Gap at provincial level

The choropleth maps referred to PGEI estimate for Italy (Fig. 9), highlight a gradually worsening situation, moving North to South of the country. This situation also occurs for each domain of the index, but with clear territorial differences within the macro

[20] Neighborhood information have been obtained following the queen's neighborhood association, that produced a weight matrix W.

regions[21]. Even within each region we can notice some differences[22]. In this respect, Tab. 7 shows the decomposition considering the between group variance (the reported percentage) and the within group variance (its complementary). This can be useful to understand how much information should be investigated going beyond the simple regional borders. Between group variation accounts for maximum the 86% of variability in the index and its domains, meaning that at least 14% of variability should be investigated more locally than at NUTS2 level. This is especially true for health and power domains, where the NUTS2 aggregation accounts respectively for 58% and 59% of the entire variability.

A deeper analysis can be conducted by each domain (Fig. 10). Minimum and maximum values for health domain are very high, also due to the use of a single basic indicator (*health_1.1*). The best situation seems to be observed for Central Italy and the North-East, while both the North-West and the South macro regions show darker peaks corresponding to lower parity levels.

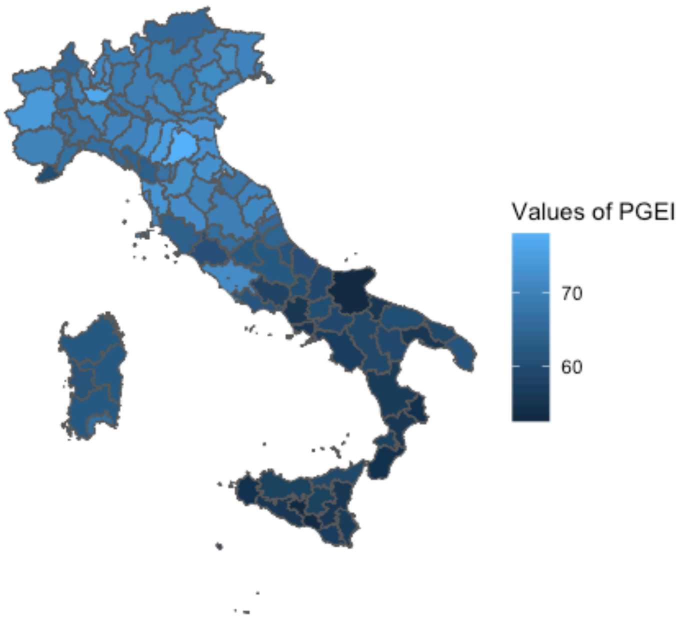


Figure 9: Provincial Gender Equality Index by province (PGEI estimates, 2018)

Domain	Between group variances using NUTS2 classification
PGEI	86%
WORK	80%
HEALTH	58%
KNOWLEDGE	81%
POWER	59%
MONEY	84%

Table 7: Between group variance explanatory power.

[21] NUTS1 classification divides Italy into 5 macro-regions: North-West, North-East, Centre, South, Islands.
[22] This happens for example in Toscana and Puglia, where some provinces (Grosseto and Foggia) tend to have worse results than others (Pisa and Bari).

Money domain shows a better situation in Northern Italy. Areas of Emilia Romagna, Piemonte and Friuli Venezia Giulia seem to show the best results. Still, within these three regions some provincial differences persist. In this domain, provinces showing a worse situation are mainly located in Sicilia and Puglia. Also in this case, differences in results between provinces can be detected.

Knowledge domain shows a situation with fewer differences between North and South, with more differences between mainland Italy and the islands (Sicilia and Sardegna). The area of Southern Sardegna shows the most important criticalities, with strong disparity in knowledge domain and a domain value around 50. Work domain, considering all domains, is the one that seems closer to PGEI trends, showing a situation that gradually worsens from North to South. Even in this case, some regional heterogeneity can be found (e.g., in Emilia Romagna, Sicilia, Sardegna).

Analyzing the domain of Power, it is necessary to consider the scale, which is lower than all the others, as the maximum value for power is set around 50 points. Compared to the other domains and to the PGEI results, the territorial dynamics are much more heterogeneous than the simple North vs South dichotomy. For example, in Emilia-Romagna we find the best level of the index, as well as some areas of Piemonte, Veneto, Umbria, and Sardegna. The worst situation seems to be found in Calabria, Puglia, Basilicata and Marche. Also in this case, different results are observed within the same regions. The most striking, in this case, are the results of Emilia Romagna, Piemonte, Puglia, and Sicilia, showing very different color compositions.

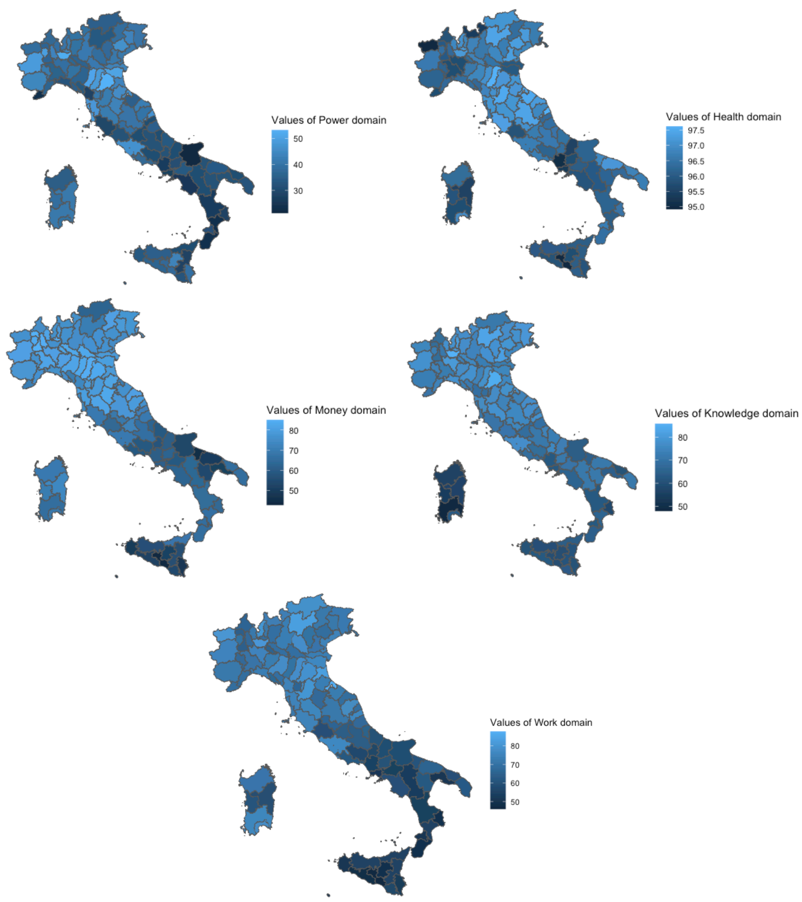


Figure 10: PGEI domains estimates by province (2018)

As shown in Fig. 11, Local Moran index values vary from negative values (with a minimum of -1) to positive values (the maximum, +3). Recall that the Global Moran Index is included between 0.6 to 0.8.

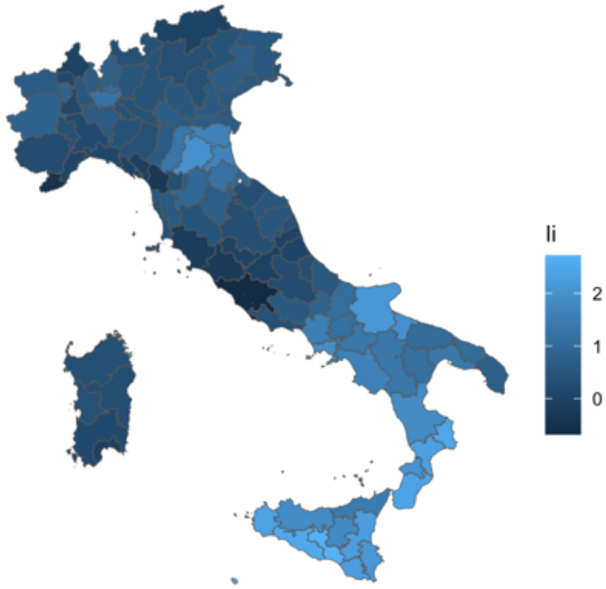


Figure 11: Local Moran's index by province

Without considering the significance level of a LISA indicator (Fig. 12a), we can divide Italy into two macro clusters with positive correlations: a "high PGEI" cluster, where provinces show an above-the-average PGEI[23]. This cluster is mainly located in the Northern/Central regions. A "low PGEI" cluster for the Central / Southern regions and the islands. Some outlier provinces can be found both in the Northern / Central regions (i.e., Piacenza, Forlì-Cesena, Macerata) and in Southern Italy (i.e., Matera, Catanzaro). All these regions are characterized by an inverse correlation of their equality level with respect to their neighbors one. Their counter tendency, if significant, can suggest a local instability of those areas when compared to the national trend (Loonis, 2018).

However, these outliers are no more consistent if we introduce a significant level threshold[24]. Several provincial clusters are significant, both in Northern / Central Italy (with a higher level of PGEI) and in Southern / Insular Italy (with a lower level of PGEI). At alpha = 0.05[25], only clusters located in Toscana, Veneto, Campania, Lazio, Abruzzo e Sicilia are still valid. Further reducing the alpha level (0.01)[26], the negative cluster in Southern Italy remains unchanged, whereas the only remaining Central cluster is observed in Toscana.

[23] Greater than the average Italian PGEI value, set to 65.47.
[24] Alpha = 0.1 presented in figure 12.b.
[25] See figure 12.c.
[26] See figure 12.d.

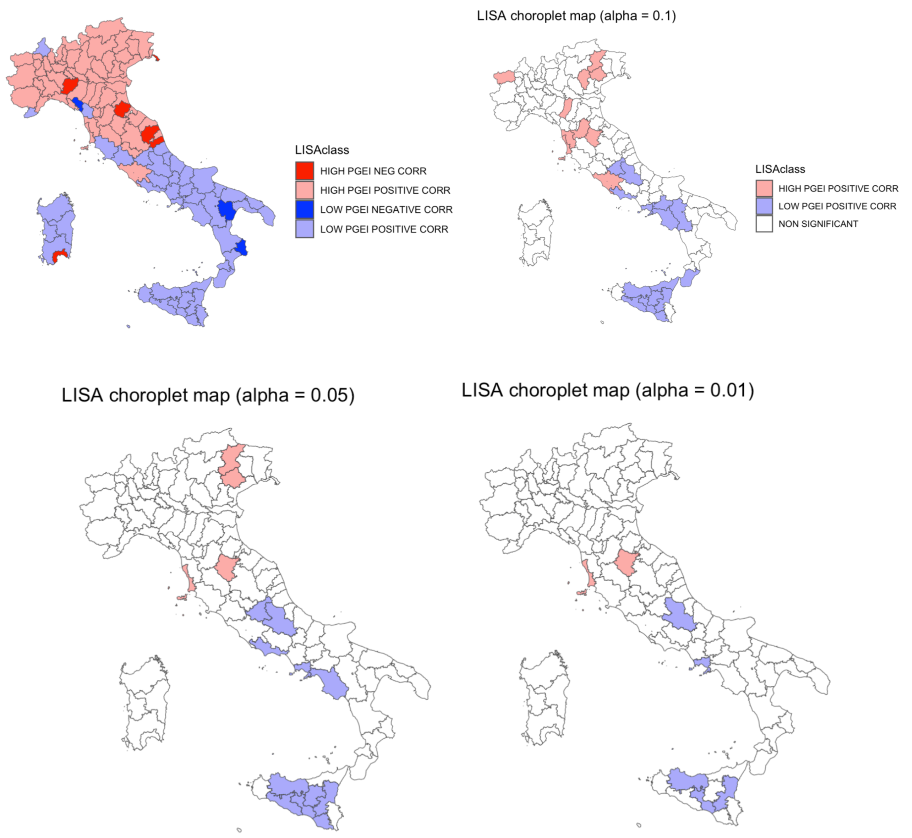


Figure 12: Local Moran's classification, by province and significance level. Different values of alpha are presented (12.B: alpha= 0.1; 12.C: alpha = 0.05; 12.D: alpha = 0.01), including a general overview without considering any significance level (12.A)

Findings suggest that the Central nucleus of higher PGEI provinces includes the provinces of Livorno and Arezzo. The provinces of Napoli, L'Aquila, and several Sicilian provinces draw other Southern and Insular nuclei characterized by lower PGEI levels. In all of them, the surrounding provinces maintain an acceptable significance linkage, identifying the neighbors of the nucleus. Interestingly, the Sicilian cluster extends beyond both NUTS2 and geographical borders, including a province that belongs to a different region, located in the mainland (i.e., Reggio Calabria).

As underlined in section 2.5, to better understand the neighborhood extension, we should also consider the non-significant neighbors of a significant nucleus. Looking at the Centre cluster, once we consider the surrounding provinces, Livorno and Arezzo became part of the same cluster located within the same regional borders (Toscana). On the other side, the two Southern nuclei (Napoli and L'Aquila) do not share any contiguous province with significant correlations in their cluster, keeping their neighborhood areas separated one from the other. In particular, the province of L'Aquila seems to be correlated with the province of Isernia, which belongs to a different NUTS2 region (Molise). Finally, the neighborhood extension for the Sicilian nucleus includes the entire provinces included in the region as a single nucleus. Especially the Central nucleus suggest that a cluster analysis applied beyond NUTS2 borders has its own rationale, that would not be highlighted if the analysis will be limited to a macro point of view.

4. Conclusions

Our methodology proposes a consistent index for measuring gender equality at the Italian local level (NUTS3), so far missing in the field literature. As we adapted an already existing index, we took into consideration several source of uncertainty. However, the proposed gold standard methodology produced consistent results across discarded methodologies, increasing the validity of findings. Estimations were able to detect a spatial correlation for gender discrimination. This statistically significant correlation is valid both at a global level and the local level. With a national perspective, Italian provinces tend to be positively correlated between each other. At a local level, we were able to identify four main clusters positively correlated within their neighborhoods have been identified: a cluster located in Central Italy, with a gender equality index above the national average; three clusters located in Central / Insular Italy, with gender equality levels below the national average. In particular, the Central cluster of Isernia and L'Aquila includes provinces of different NUTS2 regions.

Our analysis showed that there is space for a local focus. Not only the original dichotomy between Norther and Southern Italy is limited, but also intra-regional heterogeneity can be found with respect to the phenomenon. We encourage to follow this research pattern, despite we should be aware that using local data can be a threaten to usual data independency assumption, because local correlation can be found in some Italian areas, as highlighted by our results. This local correlation is also a signal that administrative boundaries (i.e NUTS classification) are not fully explicative of social patterns, as economics and social phenomenon can be correlated across different regions.

Further developments can be implemented to the index version we propose, as we were not able to fully adapt the original index. This problem is mainly linked to a missing data availability at a more local level. In our case, we were not able to totally adapt a single index dimension (use of time). Still, results suggests that differences arose from this missing dimension are minimal when the full index is estimated. Promoting within National Statistical Institutes a data collection with a greater local significance can be too expensive and not immediately feasible. Still, researcher can try to fill this data gap and complete the full adaptation procedure. As an example, a further approach can be the one of exploiting small area methods (Pratesi *et al.,* 2020; Dadoug *et al.,* 2023).

As final remark, we should consider that this powerful index is affected by some limitations. The first is due to the conceptualization of "gender", that here is expressed just throughout a binary classification (females vs males). As a matter of fact, the nature of gender possibilities is recognized as a wider spectrum, including several non-binary categories (e.g., being recognized as a trans person). Still, Official Statistics data are well-behind the collection of data at such a disaggregated level, and the academic community is still trying to understand which difficulties implies collecting information on Sexual Orientation and Gender Identity (SOGI) (Holzberg *et al*., 2019). A further limitation is caused by the intersection of categories that potentially produced different sources of discriminations (i.e., being an immigrated female, being a female citizen with disabilities, …). In this respect, EIGE developed a satellite domain called intersectional inequalities that uses disaggregated data for several combination of categories (i.e., gender, level of education, country of birth, …). A clear drawback is that, especially at more local levels of analysis (i.e., NUTS3), this disaggregation is not always taken into account in the survey methodology.

Acknowledgment

This research project has been supported by the 60% University funds (University of Bergamo, 2021 and 2022). Part of this work has been developed in the framework of an internship supervised by the Italian Official Statistical Institution (ISTAT), with the supervision of Sergio Pollutri and of Prof. Federica Maria Origo.

Disclaimer: the contents of this article reflect the views of the authors and not necessarily the official views or opinions of ISTAT.

References

Anselin, L. (1995), Local Indicators of Spatial Association—LISA. Geographical Analysis, 27(2), 93-115.

Anselin L. [GeoDa Software] (2016). Local Spatial Autocorrelation [video]. Youtube link at: https://youtu.be/yMcuYRJo1CM?si=7pX8n_auehxzVezd

Audric, S., M., De Bellefon, E., Duriex. Descriptive Spatial Analysis. Chapter in in Loonis & De Bellefon (2018).

Badgett, M.V.L., Hartmann, H.L. (1995). The effectiveness of equal employment opportunity policies in Simms, M.C. (ed.), Economics Perspectives on Affirmative Action, University Press of America, Maryland.

Bakker, M., Torbeyns, J., Wijns, N., Verschaffel, L., De Smedt, B. (2018). Gender equality in 4- to 5- year-old preschoolers' early numerical competencies. Developmental Science, 22(1), 1-7.

Bivard R. (2022). R Packages for Analyzing Spatial Data: A Comparative Case Study with Areal Data. Geographical Analysis, 54(3), 488-518.

Borgonovi F., Choi A., Paccagnella M. (2021). The evolution of gender gaps in numeracy and literacy between childhood and young adulthood. Economics of Education Review, 82, 1-18.

Branisa, B., S. Klasen and M. Ziegler (2009), The construction of the Social Institutions and Gender Index (SIGI), Background Paper, OECD, Paris.

Branisa, B., S. Klasen, M. Ziegler, D. Drechsler and J., Jutting (2014). The Institutional Basis of Gender Inequality: The Social Institutions and Gender Index (SIGI). Feminist Economics, 20(2), 29-64.

Cascella, C., Williams, J., Pampaka, M. (2022). An Extended Regional Gender Gaps Index (eRGGI): Comparative Measurement of Gender Equality at Different Levels of Regionality. Social Indicators Research, 159(2), 757–800

Cederman, L.E., Weidmann, N.B., Gleditsch, K.S. (2011). Horizontal inequalities and ethno- nationalist civil war: a global comparison. The American Political Science Review, 105(3), 478-495.

Constantin A., Voicu M. (2015). Attitudes Towards Gender Roles in Cross-Cultural Surveys: Content Validity and Cross-Cultural Measurement Invariance. Social Indicators Research, 123(3), 733–751.

Dagdoug, M., Goga, C., Haziza, D. (2023). Model-assisted estimation in high-dimensional settings for survey data. Journal of Applied Statistics, 50(3), 761-785.

Daly, M. (2005). Gender Mainstreaming in Theory and Practice. Social Politics: International Studies in Gender, State & Society, 12(3), 433-450.

De Bellefon, E., Le Gleut, R. and V. Loonis. Codifying the neighbourhood structure. Chapter in in Loonis & De Bellefon (2018).

Di Bella E., Leporatti L., Gandullia L., Maggino F. (2020). Proposing a regional gender equality index (R-GEI) with an application to Italy. Regional Studies, 44(5), 962-973.
Duncan, S. (1995). Theorizing European gender systems. Journal of European Social Policy, 5(4), 263–284

EIGE–- European Insititute for Gender Equality (2013). Gender Equality Index report. Publications Office of the European Union, Luxembourg.

EIGE–- European Insititute for Gender Equality (2017). Gender Equality Index report. Publications Office of the European Union, Luxembourg

EIGE–- European Insititute for Gender Equality (2022). Gender Equality Index 2022: The COVID-19 pandemic and care. Publications Office of the European Union, Luxembourg

Ferrant, G., Fuiret, L., Zambrano, E. (2020). The Social Institutions and Gender Index (SIGI) 2019: A revised framework for better advocacy. Working Paper N. 342. OECD Development Centre, Paris.

Hausmann, R., Tyson, L., & Zahidi, S. (2006). *The Gender Gap Index 2006: a new framework for measuring equality. The Global Gender Gap Report*. World Economic Forum, Geneva

Hausmann, R., Tyson, L.T., Zahidi, S. (2009). The Global Gender gap report. World Economic Forum, Geneva.

Hausmann, R., Tyson, L.T., Zahidi, S. (2012). The Global Gender gap report. World Economic Forum, Geneva

Holzberg, J., Ellis, R., Kaplan, R., Virgile, M., Edgar, J. (2019). Can They and Will They? Exploring Proxy Response of Sexual Orientation and Gender Identity in the Current Population Survey. Journal of Official Statistics, 35(4), 885-911.

ISTAT (2019). Le differenze territoriali di benessere. Una lettura a livello provinciale. ISTAT, Rome

Lomazzi, V., Crespi, I. (2019). Gender mainstreaming and gender equality in Europe: Policies, culture and public opinion (1st ed.). Bristol University Press. Bristol.

Loonis, V., De Bellefon, M. (2018). Handbook of Spatial Analysis. Theory and practical application with R. INSEE / Eurostat, Montrouge Cedex

Mazziotta M., Pareto A. (2020). Gli Indici Sintetici. G. Giappichelli Editore, Torino

OECD (2008). Handbook on Constructing Composite Indicators. Methodology and user guide. OECD Publishing, Paris.

Richardt, N. (2008). Gender Gap in Darity JR. W. A. (curated by), International Encyclopedia of the Social Sciences (2nd edition, Vol. 3). Macmillan Social Science Library, USA, 77-279

Salima, B. and M. De Bellefon. Spatial Autocorrelation Indices. Chapter in in Loonis & De Bellefon (2018).

Schultz Lee, K.S., Tufiş, P.A. Alwin, D.F. (2010). Separate spheres or increasing equality? Changing gender beliefs in postwar Japan. Journal of Marriage and Family, 72(1), 184–201.

Stewart, F. (2008), Horizontal Inequalities and Conflict: Understanding Group Violence in Multiethnic Societies, Palgrave, London.

Permanyer, I. (2015). Why call it “equality” when it should be “achievement”? A proposal to uncorrect the “corrected gender gap” in the EU Gender Equality Index. Journal of European Social Policy, 25(4), 414-430

Pratesi, M., Quattrociocchi, L., Bertarelli, G., Gemignani, A., Giusti, C. (2020). Spatial Distribution of Multidimensional Educational Poverty in Italy using Small Area Estimation. Social Indicator Research, 156(2-3), 563-586

Stiglitz J., Fitoussi J., Durand M. (editors) (2018). For Good Measure: Advancing Research on Well- being Metrics Beyond GDP. OECD Publishing, Paris.

UNDP – United Nation Development Programme (2022). The 2021/2022 Human Development Report. UNDP, New York.

Verloo, M. (2005). Displacement and Empowerment: Reflections on the Concept and Practice of the Council of Europe Approach to Gender Mainstreaming and Gender Equality. Social Politics: International Studies in Gender, State and Society, 12(3), 344–365.

Walby, S. (2005). Gender Mainstreaming: Productive Tensions in Theory and Practice. Social Politics: International Studies in Gender, State and Society, 12(3), 321–343.

Statistical annexes

Annex 1: Principal Components Analysis for work domain

Mean and Variance

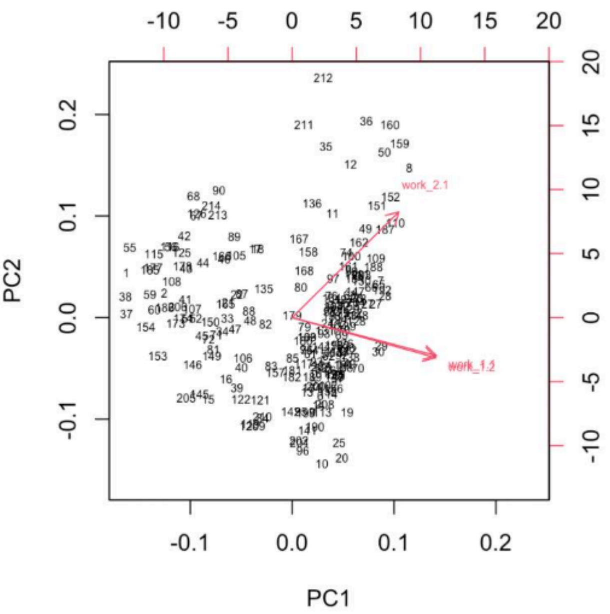
stat	work_1.1	work_1.2	work_2.1
mean	75.07	86.58	54.52
variance	170.55	92.52	114.49

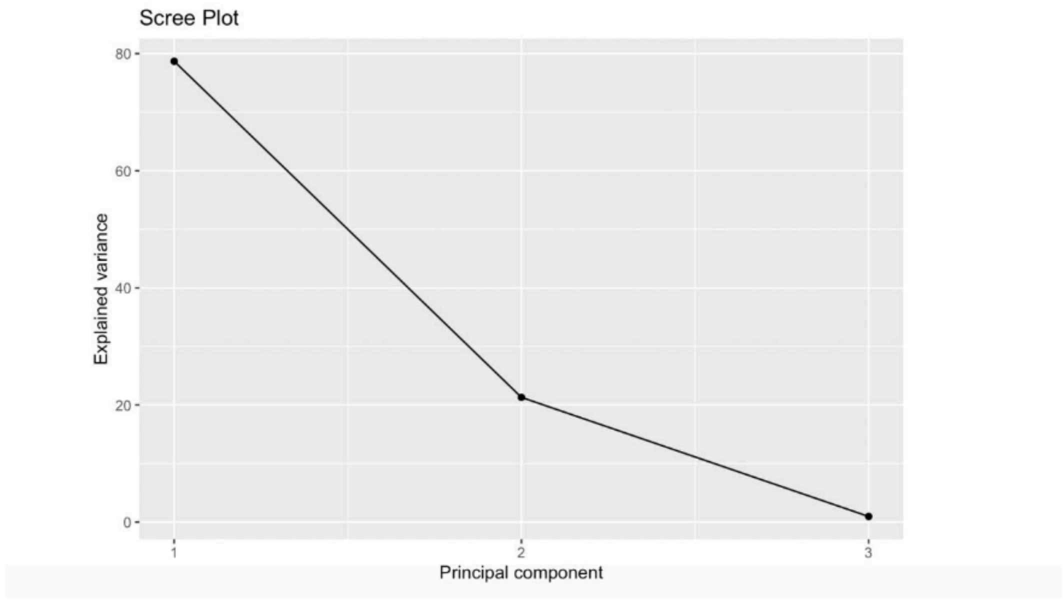
Explained variance

var %	PCA
78.7	PC1
21.3	PC2

PCA components

	PC1	PC2
work_1.1	0.63	-0.32
work_1.2	0.63	-0.34
work_2.1	0.46	0.89





Annex 2: Principal Components Analysis for knowledge domain

Mean and Variance

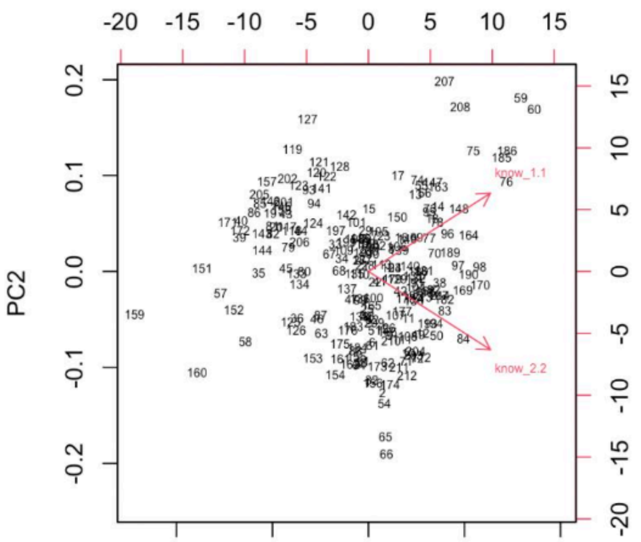
stat	know_1.1	know_2.2
mean	65.35	79.31
variance	40.11	108.41

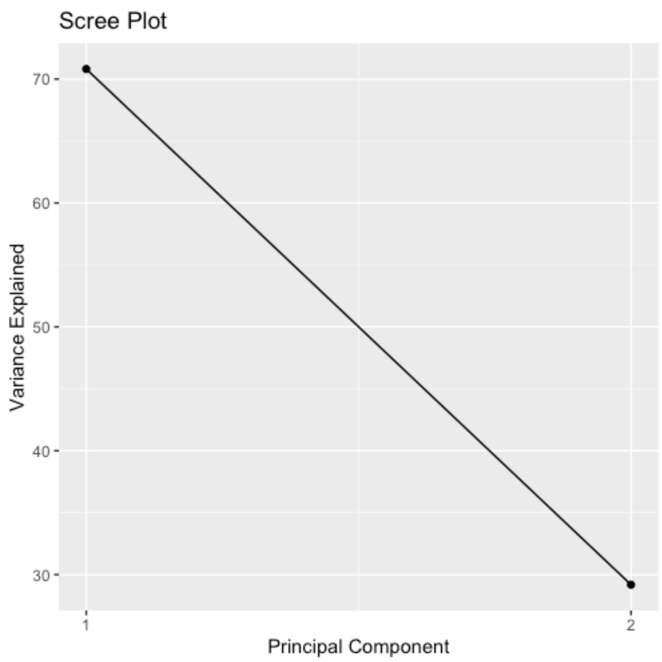
Explained variance

var %	PCA
70.81	PC1
29.19	PC2

PCA components

	PC1	PC2
know_1.1	0.71	0.71
know_2.2	0.71	-0.71





Annex 3: Principal Components Analysis for money domain

Mean and Variance

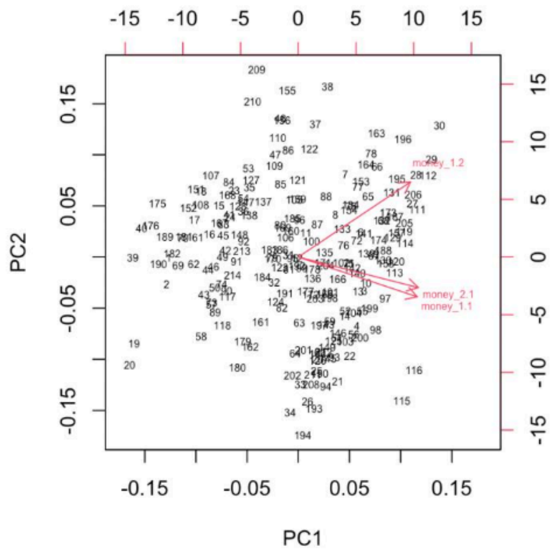
stat	money_1.1	money_1.2	money_2.1
mean	63.93	75.01	72.57
variance	47.77	8.39	293.65

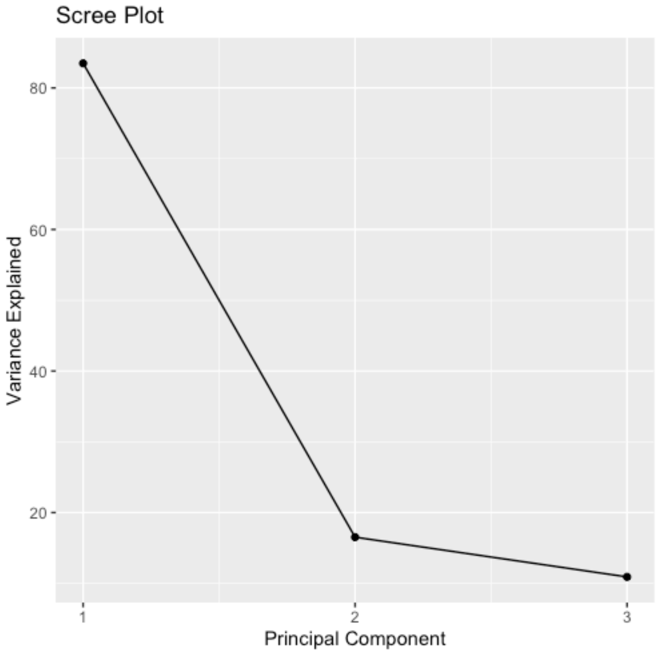
Explained variance

var %	PCA
83.47	PC1
16.53	PC2

PCA components

	PC1	PC2
money_1.1	0.59	-0.44
money_1.2	0.55	0.83
money_2.1	0.59	-0.34





Annex 4: Principal Components Analysis for power domain

Mean and Variance

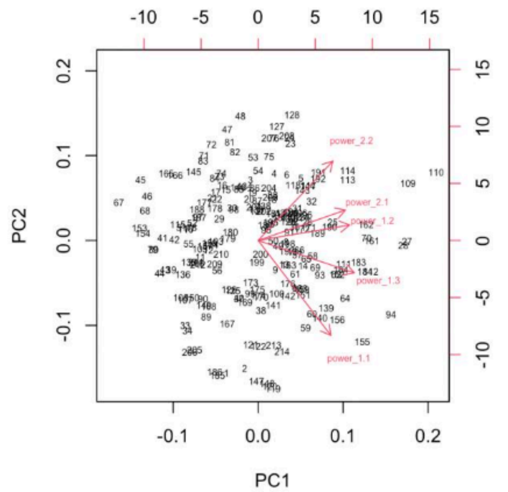
stat	power_1.1	power_1.2	power_1.3	power_2.1	power_2.2
mean	42.58	27.71	65.12	30.07	30.06
variance	512.87	293.96	72.85	88.55	25.91

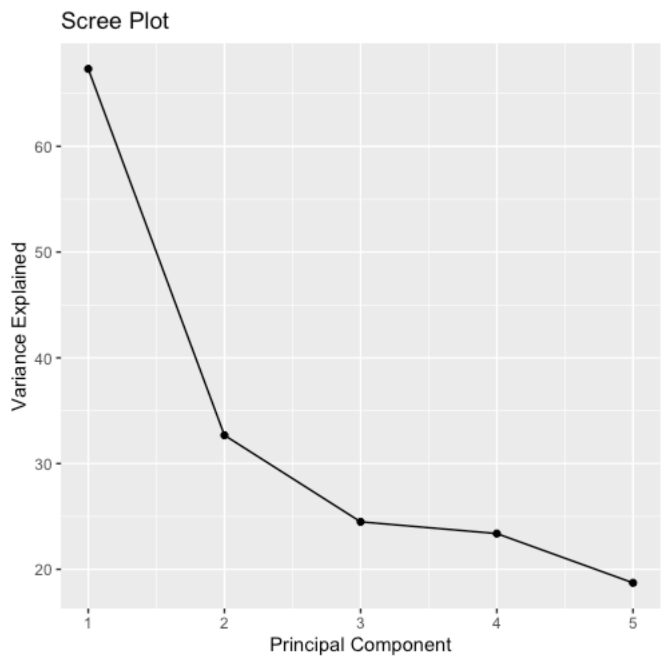
Explained variance

var %	PCA
54.08	PC1
26.25	PC2
19.67	PC3

PCA components

	PC1	PC2	PC3
power_1.1	0.38	-0.72	-0.24
power_1.2	0.48	0.12	-0.30
power_1.3	0.51	-0.24	0.13
power_2.1	0.46	0.23	0.78
power_2.2	0.39	0.6	-0.47





	PGEI	health	know	money	power	work	health_1	know_1	know_2	money_1	money_2	power_1	power_2	work_1	work_2
PGEI	1.000														
health	0.549***	1.000													
know	0.828***	0.545***	1.000												
money	0.889***	0.404***	0.665***	1.000											
power	0.746***	0.362***	0.434***	0.553***	1.000										
work	0.924***	0.507***	0.735***	0.790***	0.564***	1.000									
health_1	0.549***	1.000***	0.545***	0.404***	0.362***	0.507***	1.000								
know_1	0.509***	0.431***	0.748***	0.350***	0.228***	0.430***	0.431***	1.000							
know_2	0.824***	0.485***	0.915***	0.698***	0.456***	0.746***	0.485***	0.416***	1.000						
money_1	0.868***	0.414***	0.697***	0.884***	0.606***	0.759***	0.414***	0.509***	0.645***	1.000					
money_2	0.818***	0.361***	0.586***	0.969***	0.475***	0.733***	0.361***	0.233***	0.661***	0.740***	1.000				
power_1	0.639***	0.313***	0.287***	0.450***	0.971***	0.476***	0.313***	0.100	0.332***	0.465***	0.400***	1.000			
power_2	0.716***	0.339***	0.708***	0.616***	0.575***	0.566***	0.339***	0.545***	0.638***	0.767***	0.480***	0.362***	1.000		
work_1	0.941***	0.485***	0.764***	0.866***	0.586***	0.949***	0.485***	0.395***	0.807***	0.803***	0.820***	0.488***	0.614***	1.000	
work_2	0.555***	0.370***	0.410***	0.346***	0.313***	0.748***	0.370***	0.348***	0.350***	0.395***	0.288***	0.280***	0.261***	0.500***	1.000

Annex 5: Pearson correlation index for PGEI, domains and subdomains