Trust and security in Italy

Silvia Golia

1. Introduction

Starting from 2018, the European University Institute (EUI) and the corporate YouGov implemented a survey aimed to study the evolution of European, transnational solidarity denoted as EUI-YouGov survey on Solidarity in Europe (Hemerijck et al., 2021). At the moment four waves (2018, 2019, 2020, 2021) are available for the analysis. The survey covers many aspects of the solidarity (issues, instruments and beneficiaries of the solidarity) plus other dimensions related to it, such as security and trust in the own government or in the European Union (EU). The survey was administered to a representative sample of citizens from 11 (2018) to 13 (2021) EU member states plus the United Kingdom, and was carried out online during the month of April. The datasets are freely available for download [1].

The sections of the questionnaire evolved during the four waves, adding new questions, revising the text of some of the old questions and eliminating some other questions. Nevertheless, there were sections remained unchanged over the years, such as the ones concerning security and trust in national government and EU. The interesting thing in these three sections is that they are composed of the same 10 areas.

The data are not longitudinal, given that the subjects change at each time span, so the four waves can be considered together. This paper starts from this characteristic to investigate if and how the feeling of security and trust about the 10 areas changes over the time, and the tool used is the Differential Item Functioning (DIF) analysis across time. DIF analysis was born as a tool to assess the validity of a scale, given that it tests the invariance of an item with respect to the characteristics of the subjects (a typical example is the gender); if an item shows DIF then, in most cases, it has to be revised or deleted. Instead, in this paper the primary interest is to study the possible evolution of the items difficulty in order to get insights on what the population felt in these four years.

Moreover, given that the period of study is 2018-2021 and the administration was done in April, the answers of the first two years refer to a Covid-19 pre-pandemic period, whereas the answers collected in the following two years are referred to the pandemic period, and this is another interesting aspect of these data.

The paper is organized as follows. Section 2 reports a brief description of the tools used in the analysis whereas Section 3 the description of the main findings. Conclusions follow in Section 4.

2. Methods

The model used in the paper to take into account the available data and to hit the aim of the study is the Rating Scale Model (RSM) (Andrich, 1978), which belongs to the family of the Rasch models. RSM turns raw scores into linear and reproducible measures expressed in logits. Given an item i with $m+1$ response categories ($c = 0, 1, \cdots, m$), according to RSM the probability of the subject s with level of latent trait θ_s (denoted also as the ability of the subject

[1] https://cadmus.eui.eu/handle/1814/72778

Silvia Golia, University of Brescia, Italy, silvia.golia@unibs.it, 0000-0003-0015-8126

Referee List (DOI 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

s) to respond in category c is given by:

$$P(X_{si} = c) = \frac{exp\left\{c(\theta_s - \delta_i) - \sum_{j=0}^{c} \tau_j\right\}}{\sum_{l=0}^{m} exp\left\{l(\theta_s - \delta_i) - \sum_{j=0}^{l} \tau_j)\right\}} \quad (1)$$

where δ_i represents the difficulty of item i and the τ_j are called thresholds ($\tau_0 \equiv 0$ and $\sum_{j=1}^{m} \tau_j = 0$) and they are equal for all the items. The choice to use the RSM as the measurement model, instead of other alternatives such as, for example, the partial credit model, is motivated by the fact that, in the present study, all the items forming each questionnaire make use of the same response format. Therefore, it is reasonable to assume that the test constructors, respondents, and test users all perceive the items to share the same rating scale (Linacre, 2000). All the parameters are expressed in the same scale (logit) and this allows comparisons. The difficulties of the items δ can be compared between each other and also with the abilities distribution. The estimates of all the parameters involved in RSM are done imposing that $\sum_{i=1}^{k} \delta_i = 0$, where k is the number of items, and this implies that zero is the average item difficulty. Items with estimated difficulty below zero are easier items, that is they are items for which it is not so difficult for the respondents to score high.

DIF refers to the different functioning of a test item for comparable groups of respondents and it is formally defined as follows. An item exhibits DIF if respondents of equal ability on the construct intended to be measured by a test, but from separate subgroups of the population, differ in their expected score on that item (Roussos and Stout, 2004). The reasons for which an item exhibits DIF are various and linked to the context of analysis. With respect to the year of interview as subject characteristic, the hypothesis about the reasons for a different functioning is the change of the external conditions from one year to the next due to the presence/absence of actions implemented by the national and/or European institutions.

There is a large literature regarding methods able to investigate DIF for both dichotomous and polytomous items, focusing primarily on the two-group case; less literature regards methods for the multiple-groups case. The context to which this paper belongs, is the one of multiple-groups and polytomous scored items, so it was addressed as follows. Firstly, the difficulties of all the items and the abilities of all subjects, regardless of their membership group, were estimated under the hypothesis that there are no DIF items. The resulting estimates of the items' difficulties can be interpreted as overall measures of them. Then, for the subjects of each group, the items' difficulties were estimated by applying the anchored maximum likelihood estimation, anchoring the measure of abilities of the involved subjects at the measure previously obtained. This anchoring procedure allows the resulting estimates of the difficulty parameters to be compared. For each item, the statistic computed taking the difference between the estimate for one group and the estimate from the main analysis and dividing it by its approximate standard error, was used to verify the null hypothesis "this item has the same difficulty as its average difficulty for all groups". It corresponds to the approximate Student's t-statistic test (Linacre, 2022). Moreover, the previous group DIF statistics for each item can be summarized as a chi-square statistic, which allows one to verify whether the observed DIF within each item is due to chance alone; the null hypothesis is "this item has no overall DIF across all groups" (Linacre, 2022). The test statistic is computed summing the Student's t-statistics, previously squared and normalized applying the Peizer and Pratt transformation (Peizer and Pratt, 1968).

Moreover, given that the first set of tests compares the item difficulty of one group versus the item difficulty under the hypothesis that there is no DIF, the Mantel test (Mantel, 1963) for pairwise testing for DIF was applied.

3. Results

As stated in the introduction, between the sections remained almost unchanged over the years in the EUI-YouGov survey on Solidarity in Europe, the ones concerning security and trust in the national government and the EU were analyzed in this paper. The common question regarding *Security* is "how secure or insecure do you feel about each of the following areas?", whereas the one regarding *Trust in the national government* and *Trust in the EU* is "how much do you trust ... to make things better in the following area?". The areas (items) considered by these three dimensions are listed in Table 1; when the formulation of an area is slightly different in the security section, it is reported in parenthesis in the table. For all the three dimensions and

Table 1: *List of the items of the security and trust sections of the survey. In parenthesis the formulation for the security section*

Item		Item	
1	The economic situation	2	Climate change
3	Military defence	4	Protection against (The threat from) terrorism
5	Protection against (The threat from) crime	6	Food standards
7	Employment opportunities (in your area)	8	Your own financial situation
9	Healthcare	10	Immigration

items, there were four possible response categories; Very secure (1), Fairly secure (2), Fairly insecure (3) and Very insecure (4) for *Security* and Trust a lot (1), Trust a fair amount (2), Do not trust very much (3) and Do not trust at all (4) for *Trust in the national government* and *Trust in the EU*. They form a 4 point Likert scale. There was also an other possible response category, "Don't know enough to say", but in the analysis it was treated as a missing answer. Moreover, in order to be able to use the RSM, the response categories were reversed.

The number of citizens involved in the four waves of the survey is reported in the first line of table 2. Nevertheless, not all of them responded to all the items, so, for each of the three dimensions investigated, the citizens who responded to at least 5 of the 10 items were taken into account, in order to have a sufficient amount of information to estimate the respondents' degree of security and trust. Table 2 reports their number with respect to the dimension and the wave.

Table 2: *Number of citizens who responded to at least 5 of the 10 items*

	2018	2019	2020	2021
With "Don't Know"	1065	895	2021	2028
Security	1045	873	1953	1982
Trust in the national government	1030	867	1958	1966
Trust in the EU	1025	855	1920	1930

The analysis was conducted as follows. Firstly, the chi-square statistic, at a significance level of 0.05, for testing the hypothesis that an item has no overall DIF across all groups was considered. For the dimensions *Security*, *Trust in the national government* and *Trust in the EU* the items 3 (Military defence) and 7 (Employment opportunities in your area), the items 3 and 4 (Protection against terrorism) and the item 5 (Protection against crime) were, respectively, the only ones which did not suffer for DIF, which means that their difficulty remained stable across the years. It has to be noted that, even if the overall test rejected the null for item 3 and dimension *Trust in the EU*, the analysis of a series of pairwise Mantel tests revealed that the hypothesis of no DIF item was always accepted, so it is possible to conclude that military defence is the unique item stable across the years in common between the three dimensions. The other items did not remain invariant over the years and their trend is shown in figures 1, 2

and 3, where the blue dots correspond to the items' difficulties estimated anchoring the measure of abilities, as explained in the previous section, the red dashed line the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero, which is the average item difficulty. Analyzing the three figures, it can be observed that the difficulties of

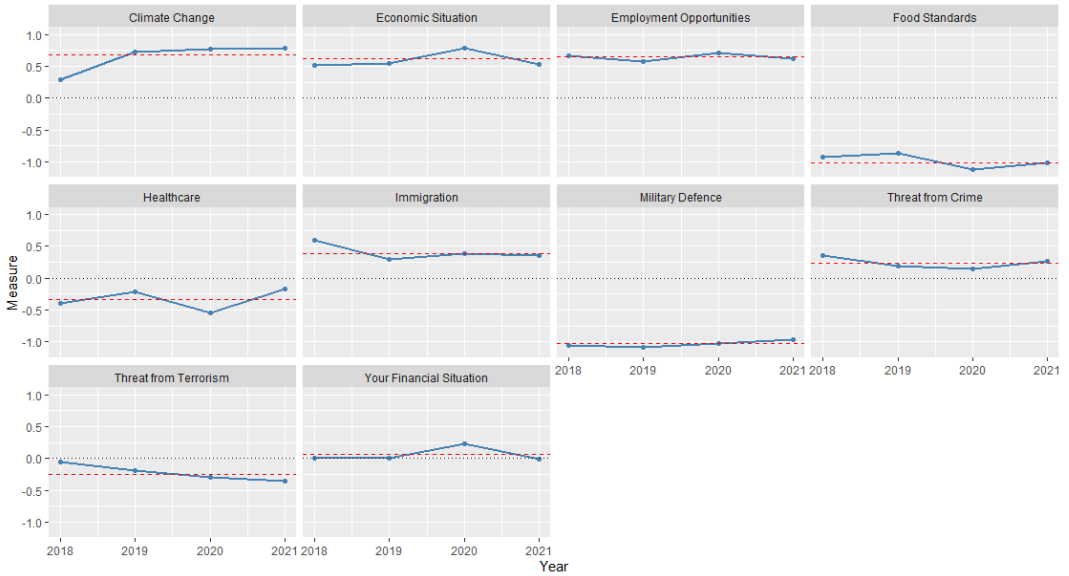

Figure 1: Security: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero

all the items exhibit a similar trend across the three dimensions, except for items 2 and 4 for which there are some differences.

Looking at item 2 (Climate Change) it can be noted that there is a jump in the difficulty to feel secure about it moving from 2018 to 2019, and then there is a light increasing trend in the last two years. Its values indicate that from 2019 climate change became one of the themes of greater insecurity among those treated, given that its difficulties remained highly above the mean. Contextually to the 2019 peak of insecurity, there is associated a peak in the scepticism that the national government can improve the existing situation with suitable politics. Nevertheless, in the following two years the item difficulty decreased, going back to the 2018 level, even if this item has still a difficulty over the mean. It is interesting to observe the different attitude of the citizens towards what the EU can do regarding this theme. During the entire period the item difficulty remained under the mean and its value at the end of the period was lower than that of 2018. The results reveal that the citizens trust the EU more than the national government in being able to make things better regarding climate change.

Considering item 4, the feeling of insecurity regarding the threat from terrorism was decreasing along the period; this theme does not represent an issue of particular concern for the citizens, in fact the item difficulty is under the mean. Contextually, there is trust that the government and the EU are able to protect the citizens against terrorism, in fact the item difficulty remained below the mean for the entire period.

It is of interest also the behaviour of item 9 (Healthcare). One can observe that there is a drop of the difficulty of the item moving from 2019 to 2020, that is before and during the first wave of the Covid-19 pandemic, for all the three dimensions, and this drop is more pronounced for *Trust in the national government*. Despite the terrible situation experienced by the Italian

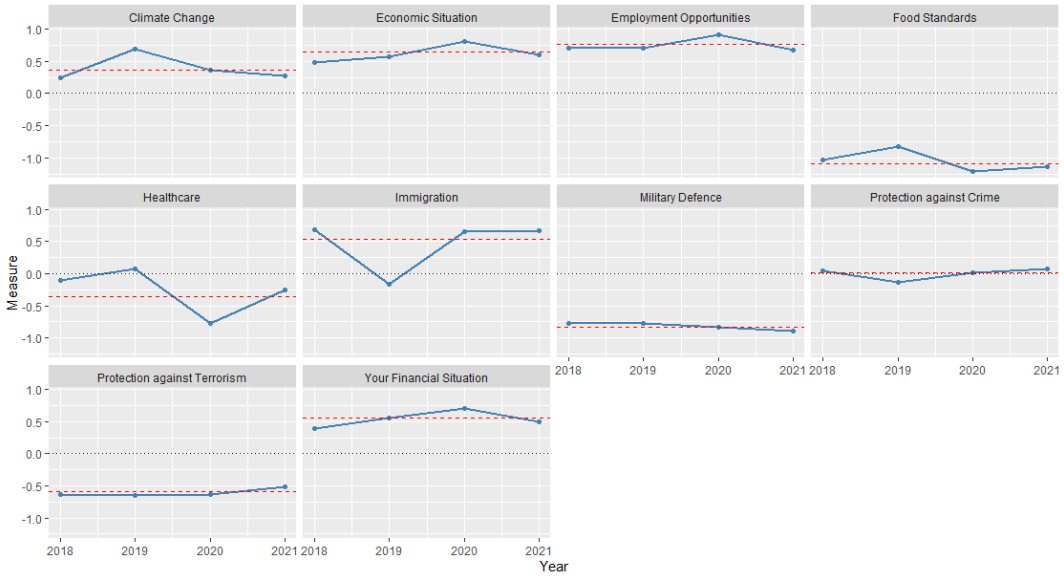

Figure 2: Trust in the national government: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dotted line highlights the zero

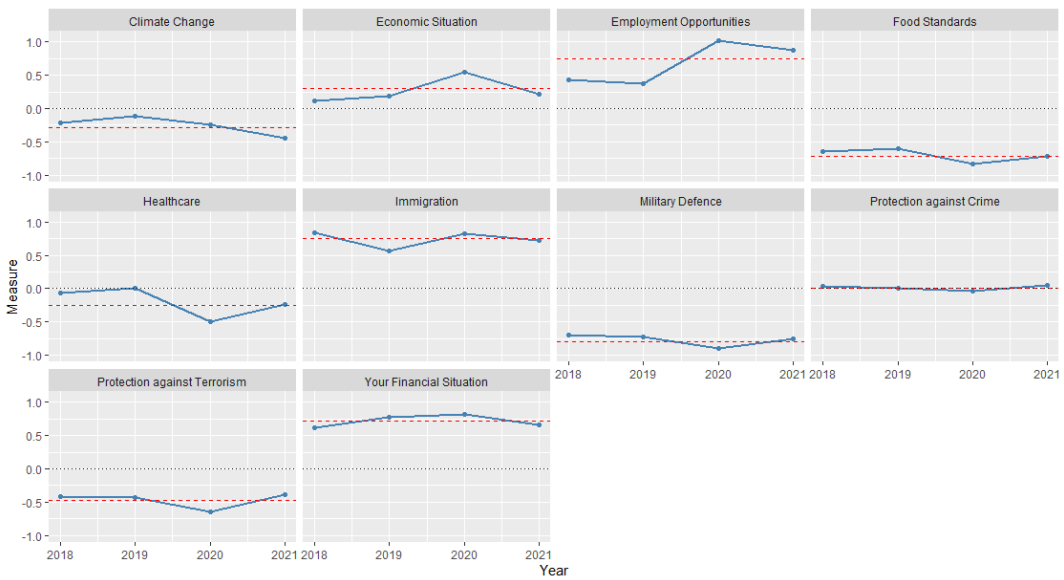

Figure 3: Trust in the EU: items' difficulties across the waves; red dashed line corresponds to the measure of the item difficulty under the hypothesis of no DIF and the dot line highlights the zero

citizens in 2020, they were confident in the actions of the Italian government and the EU to reduce the impact of the pandemic on the population. Moving from 2020 to 2021, the item difficulty increased, meaning a decrease in the citizen trust regarding the healthcare theme, even if this theme remains one of the themes of low concern, given that its difficulty remained

below the mean.

Regardless of the trend, it is of interest to highlight the different behaviour in the level of difficulty of items 5 and 8 between the three dimensions. During the entire period the citizens felt insecure regarding the threat from crime, given that the item difficulty remained above the mean, whereas they felt not sceptical that the national government or the EU could make things better regarding the protection against crime (the item difficulty was about the mean). A similar behavior, but reversed, can be observed considering the citizen own financial situation; the difficulties of the item 8 for *Trust in the national government* and *Trust in the EU* were above the mean, meaning that the citizens did not trust much both the national government and the EU in improving the existing situation, whereas the item difficulties for *Security* were around the mean, meaning that they did not feel particularly insecure regarding their own financial situation.

4. Conclusions

The paper analyzed three sections of the EUI-YouGov survey on Solidarity in Europe concerning the dimensions of security, trust in the national government and trust in the EU. All of them are related to the theme of solidarity, which is the main focus of the survey. The intent of the study was to inspect if and how the feeling of security and trust about the 10 areas (items) covered by the questionnaire changed over time analyzing the trend of the items' difficulties by means of the DIF analysis. Most of the items exhibited DIF across time and interesting patterns.

Future developments of this analysis will concern the relations between the measures of each dimension and the time and between the three dimensions.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, **43**, pp. 561–573.

Genschel, P., Hemerijck, A., Nasr, M., Russo, L. (2021). *Solidarity and trust in times of Covid-19*. EUI RSC PP, 2021/11, European Governance and Politics Programme.

Hemerijck, A., Genschel, P., Cicchi, L., Nasr, M., Russo, L. (2021). *EUI-YouGov survey on solidarity in Europe trendfile and yearly datasets (2018-2021)*. EUI Research Data, Robert Schuman Centre for Advanced Studies - https://hdl.handle.net/1814/72778

Linacre, J. M. (2000). Comparing and Choosing between "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). *Rasch Measurement Transactions*, **14** (3), pp. 768.

Linacre, J. M. (2022). *Winsteps®Rasch measurement computer program User's Guide. Version 5.2.3*. Portland, Oregon: Winsteps.com

Mantel, N. (1963). Chi-square tests with one degree of freedom; extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, **58**, pp.690–700.

Peizer, D.B., Pratt, J.W. (1968). A Normal Approximation for Binomial, F, Beta, and Other Common, Related Tail Probabilities, I. *Journal of American Statistical Association*, **63**(324), pp. 1416–1456.

Roussos, L.A., Stout, W. (2004). Differential item functioning analysis: Detecting DIF item and testing DIF hypotheses, in *The Sage handbook of quantitative methodology for the social sciences*, eds. D. Kaplan, Sage Publications, Thousand Oaks (CA), pp. 107–116.