UCLouvain

FACULTE DES SCIENCES

Faculté des sciences

Measurement Invariance of Political Efficacy

The Case of the European Social Survey

Author: **Jessica De Rongé**
Supervisors: **Christian Ritter, Pierre Baudewyns**
Reader: **Christian Hafner**
Academic year 2023-2024
Master [120] en statistique, orientation générale

«*La comparaison n'est pas raison* »

Proverb from the 13^{th} century

Summary

The European Social Survey is a cross-national survey organised in Europe every two years. The survey gives a great deal of importance to the methodological aspects of collecting high quality data with which longitudinal and cross-national comparisons can be made. One important concept in political sciences, that is studied in the ESS, is political efficacy defined as the citizen's evaluation of his competence in the political sphere as well as his evaluation of the responsiveness of the political system to his demands. Researchers have demonstrated that the political efficacy concept is linked to political participation.

Measurement invariance analysis checks if a latent variable is measured in the same way in different groups. In the ESS, the measurement invariance of the political efficacy has never been tested for the last three rounds of the questionnaire. The lack of study of this subject means that scientists who have done cross-national comparisons on political efficacy with the ESS data have made conclusions without verifying that the concept is measured similarly across the different cultures.

This research has analysed the measurement invariance characteristics of political efficacy in the rounds 8 to 10 of the ESS. Two different types of analysis have been produced. The first one has checked if it is valid to compare latent score of one country between the different rounds. The second type of analysis has consisted in checking if in one round, it is valid to compare the latent score of the different countries.

The first analysis that verified that it is correct to study the evolution over time of political efficacy in each country has shown that scalar invariance was reached for every country except Austria and Poland. This means that it is valid to compare directly differences between latent scores in the different rounds for every country except two.

The cross-national analysis by rounds has demonstrated that scalar invariance was not achieved in any of the three rounds. This means that differences observed between countries in one round should not be directly compared as part of the observed differences is probably an artificial difference caused by a bias in the questionnaire.

Finally, the lack of scalar invariance has been explored. The invariance has been represented graphically to let the reader evaluate the level of non-invariance that was present in the data for each round. The lack of invariance was small. By looking at the invariance graph and the latent means graph, it was possible to see that the invariance between countries was different between sub-groups of countries and that some sub-groups could have achieved scalar invariance.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisors, Professor Ritter and Professor Baudewyns for their help in the completion of my master thesis. Your guidance, advices, feedback and suggestions were invaluable to improve my research.

I would also like to thank Aurélie Bertrand who made herself available to answer some of my questions regarding the methodology.

Next, I would like to extend my heartfelt thanks to my family, especially my mother, my father and my brother who were an invaluable source of support throughout the completion of my master in statistics. Thank you for your encouragement and your belief in me, even when I was casting doubt about my ability to conduct this project until the end. Thank you for your careful proofreading even when the topic was out of your comfort zone.

Finally, I would like to acknowledge my friends who were always there when I needed a break from my master thesis.

Acronyms

AL Albania

AT Austria

BE Belgium

BG Bulgaria

CH Switzerland

CY Cyprus

CZ Czechia

DE Germany

DK Denmark

EE Estonia

ES Spain

FI Finland

FR France

GB United Kingdom

GR Greece

HR Croatia

HU Hungary

IE Ireland

IS Iceland

IL Israel

IT Italy

LT Lithuania

LV Latvia

ME Montenegro

MK North Macedonia

NL Netherlands

NO Norway

PL Poland

PT Portugal

RO Romania

RS Serbia

RU Russian Federation

SE Sweden

SI Slovenia

SK Slovakia

PCA Principal component analysis

CFA Confirmatory factorial analysis

ML Maximum likelihood

WLS Weighted least square

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Most European countries have a political system that is based on liberal democracy in which one of the major forms of political participation is election. However, since 1990, the election turnout in the different countries has been declining. The most surprising is that the decline is much more present in new democracies (Solijonov, 2016). Different researchers are trying to understand this trend, especially as it is a phenomenon that affects almost all European countries. Researchers are currently studying the reasons behind the lack of involvement in elections as well as understanding the development of alternative forms of participation. One key factor used to study political participation is political efficacy.

In Europe, one of the most well-known cross-national surveys about social sciences is the European Social Study (ESS). This survey collects information notably about political efficacy in different countries of Europe. This makes the ESS a very useful tool to produce cross-national comparisons. However, this type of survey needs to be carefully implemented to obtain data that can be compared between countries without bias.

Cross-cultural comparison can be made only if the instrument that collects the data in the different cultures can be characterised as equivalent in the different contexts. This is also known as measurement invariance. It is the guarantee that the data collected is culturally unbiased and permits the direct comparison between countries. Measurement invariance is essential when the concept under study is not directly observable and that researchers rely on a scale composed of different items to measure the concept.

Before studying the difference between cultures, measurement invariance should first be assessed. However, in most cases, the equivalence of the concept is considered acquired if the same questionnaire is used to collect the data. This is not enough to consider the validity of the instrument cross-culturally. Measurement invariance demands specific analysis to determine the level of equivalence, but not all research needs the highest level of equivalence. Indeed, it varies depending on the comparison that the researcher intends to investigate.

Political efficacy is a latent construct. In the ESS, it is measured by four different questions in the survey. However, the ESS has not produced documentation on the measurement equivalence of the concept. Still some published work in political sciences has used the ESS data to compare political efficacy across countries. As the invariance has not been tested, it is possible to cast doubt upon the validity of their results. Indeed, when non-invariant, the difference found can be artificial difference that do not represent the true phenomena under study. These reasons explain why this work will propose an analysis of the measurement invariant characteristics of the political efficacy scale of the ESS.

This master thesis is structured as follows. The first part will present the state of the art. This section will define in more detail what is political efficacy. Once it is done, the measurement invariance concept will be presented as well as possible bias in an instrument. Next, a few words about the ESS will be written as well as the development of the political efficacy scale in the ESS. Finally, the research question will be developed.

The next section is devoted to the methodology. This section will start by presenting the different methods mobilised to answer the research question, especially the confirmatory factorial analysis. Finally, this section will introduce the data used in the analysis.

The third section of this work is dedicated to the analysis. This part will describe the results of the descriptive analysis before exploring the results of the measurement invariance analysis. The results of both longitudinal analysis (analysis for each country throughout different rounds of the survey) and the cross-national (by rounds) analysis will be presented.

The fourth section deals with the discussion of the results. Here, comparisons between the results of the different analyses are discussed and possible explanations for results indicating lack of measurement invariance are proposed.

The conclusion will summarise the most important findings of our work as well as discuss the limits of the research.

Chapter 2

State of the Art & Research Question

2.1 Political efficacy

This section is dedicated to the concept of political efficacy. Firstly, a definition of the concept is proposed, followed by a description of how political efficacy can be measured and finally, the importance of studying political efficacy will be discussed.

2.1.1 Definition of political efficacy

The political efficacy concept was first introduced by Campbell et al. (1954). They defined it as : “the feeling that individual political action does have or can have, an impact upon the political process, i.e. that it is worthwhile to perform one’s civic duties. It is the feeling that political and social change is possible, and that the individual citizen can play a part in bringing about this change” (Campbell et al., 1954, p.187). Based on this definition, political efficacy is understood as the citizen’s ability to take part in politics as well as the system ability to respond to the wishes of citizens. It is a subjective concept as it is based on the citizen’s evaluation of his own competence and on the citizen’s evaluation of the responsiveness of the political system of his country.

Since this first definition, researchers have been questioning the definition of the concept. As the first definition speaks both about citizen’s and system’s capabilities, it seems that in practice there are two different facets to the concept. Lane (1959) is one of the first authors to shed light on this intuition. Indeed, Lane believed that political efficacy should distinguish between a first dimension that would be orientated towards citizen's perception of himself, understood as his ability to take part in politics, and a second dimension that would consider citizens perception of the democratic government and political system, understood as the ability of the system to follow the will of the citizens. Balch (1974) will prove empirically that political efficacy contains two dimensions.

The first dimension of political efficacy is called "internal political efficacy". It refers to the ability to understand effectively politics as well as the competence to participate in an effective way in politics (Niemi et al., 1991). This dimension regards the citizen's evaluation of his own competences in the realm of politics.

The second dimension of political efficacy is called "external political efficacy" which alludes to the citizen's perception of the political system in which he lives. It depicts the belief of citizens regarding the response of the government to their wishes and demands as well as citizen's influence in the political system(Bene, 2020).

In practice, political efficacy should be studied as two distinct dimensions and a relation exists between the two as they both refer to the same concept. As Lane explains political efficacy "contains the tacit implication that an image of the self as effective is intimately related to the image of democratic government as responsive to the people" (Lane, 1959, p.149).

2.1.2 Measuring political efficacy

Since the 1960's, research has been conducted to find the appropriate way to measure political efficacy. One consensus is that political efficacy is a latent construct which means that it cannot be directly observed and measured. This is the case for both dimensions of the concept. Thus, to measure political efficacy, scientists have been trying to create a list of questions to measure different aspects of political efficacy and on which a scale can be obtained to measure the latent trait. Most of the work done to develop this scale has been produced on data coming from the United States. Unfortunately, no scale has achieved a consensus in the literature. One author (Morrell, 2003) has done a review of all papers using internal political efficacy and compared the different scales obtained. Most papers about political efficacy used a different questionnaire to measure the concept.

2.1.3 Political efficacy and participation

Political efficacy is an important concept for political scientists because it has been demonstrated that it has an influence on political participation. It appears that intention to vote is increased by internal political efficacy (Reichert, 2016). Furthermore, authors study the different implications of different combinations of levels of internal and external efficacy on the type of participation (Pollock, 1983). Karp & Banducci (2008) have also studied how system electoral rules can influence both political efficacy and political participation.

Those different studies are just a few that use political efficacy to explain political attitudes. The lack of consensus about a political efficacy scale questions the comparability of results across different studies. This shows that studying political efficacy scale is relevant for the advancement of the political sciences field of research.

To conclude, political efficacy is a political science concept that is measured through two distinct dimensions. There is no direct way to measure political efficacy and researchers use a questionnaire with different items to construct a scale of internal and external political efficacy. This construct is one key explanation of political participation. As political efficacy is a latent construct, it is important to verify that the instrument that collects data about this latent concept is equivalent in different contexts. If the instrument does not measure equivalently political efficacy, then any comparison of groups based on the data collected through this instrument would be biased and could lead to incorrect assumptions about political efficacy in the different groups studied.

2.2 Measurement invariance

2.2.1 Measurement invariance definition

Recently, cross-national research has experienced a growing popularity among social scientists. Cross-national research means that scientists analyse a phenomenon by comparing data from various countries/contexts. A well-known instrument to gather cross-national data is a survey. One advantage of the survey is that it permits the collection of information about a latent construct (a variable that cannot be directly measure) by using a few questions to measure different aspects of the latent construct. A scale can be derived from those questions to obtain a score on the latent construct. One point of attention in cross-national comparison is the equivalence of the data gathered by an identical instrument in the different countries. If the instrument does not accurately measure the same concept in the different contexts, then the differences in the results are not necessarily true differences between populations but artificial differences caused by a biased instrument (Widaman & Reise, 1997). In practice, most researchers assume validity when the same instrument is used in the different cultures and do not verify this assumption (Kankaraš & Moors, 2010).

Cultural equivalence of an instrument is studied by verifying the measurement invariance characteristics of the data once it has been collected. Measurement invariance is defined as “whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute” (Horn & McArdle, 1992, p. 117). The differences between the samples of data gathered by the same instrument but in different contexts are investigated to verify that observed differences come from true differences in the construct and are not artificial differences coming from a bias in the instrument. Direct comparison between cross-national data can only be made if the hypothesis of measurement invariance cannot be rejected.

In the literature, researchers have defined and conceptualised equivalence in various manners. Johnson (1998) classified equivalence’s definitions in two categories: interpretative and procedural equivalence. The first category corresponds to a conceptualisation of equivalence as the interpretation and meaning of a concept into the different cultures (Kankaraš & Moors, 2010). The second category can be defined as: “types of equivalence that are dealing with measures and procedures used in cross-cultural studies” (Kankaraš & Moors, 2010, p.122). Measurement invariance is considered as a procedural equivalence and can only be investigated if the interpretative equivalence of the concept is recognised in the different cultures studied (meaning that the construct is understood in a similar manner in the cultures studied).

2.2.2 Level of measurement invariance

In measurement invariance, different levels of equivalence can be achieved. The choice of equivalence needed for a research depends on the objective of the research and in which way, the comparison will be made. The most common types of equivalence studied in measurement invariance are configural, metric and scalar equivalence.

The first level of equivalence is configural or construct equivalence. This level of equivalence is interested in the relationship between observed and latent variables. Configural equivalence verifies that the structure between observed and latent variable is the same in the different groups under study (Steenkamp & Baumgartner, 1998; Kankaraš & Moors, 2010). In other words, configural equivalence investigates if the meaning of a latent construct is similar in different cultures. It is done by checking that the group of observed variables measuring the latent construct is the same in the different cultures (Welkenhuysen-Gybels et al., 2007). At this stage, group comparison cannot be made.

Once configural equivalence has been accepted, it is possible to look at the second level of equivalence, metric equivalence. This level of equivalence demands that in addition to a similar structure between observed and latent variables, the strength of those relationships should be equal between the different groups (Singh, 1995; Cheung & Rensvold, 2000). When an instrument achieves metric equivalence, within group comparison can be made (for example, results in different age categories of each group can be compared) as well as comparison of patterns of correlation/mean of subgroups across different groups. However, a direct comparison of latent mean between groups is incorrect if scalar invariance is not achieved (He & van de Vijver, 2012).

The last equivalence considered in this paper is scalar equivalence. It is the level of equivalence needed to be able to perform direct comparison of latent means across different groups (He & van de Vijver, 2012). Scalar equivalence builds on both configural and metric equivalence but adds another condition (to similar structure and strength of relationship). This condition demands that the latent construct has the same origin (meaning that the scale of the latent construct has the same origin in each culture studied) (Kankaraš & Moors, 2010). Equality of the origin of the measurement instrument establishes that there are no biases when computing the latent scores of the different cultural groups.

Obtaining scalar equivalence can be very difficult in some cases, especially when the number of groups is high. However, it is possible to relax the assumption in scalar equivalence to obtain partial scalar equivalence. Partial scalar equivalence consists of freeing the parameters of some observed variables. However, for cross-cultural comparison, there still needs to be a minimum of equivalent items by latent trait to be valid. Theoretically, the minimum is two, but this number limits the number of conclusions that can be made about the instrument (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000).

2.2.3 Sources of bias

The most common reason behind the lack of equivalence of an instrument in cross-cultural research is a bias (van de Vijver & Leung, 1997). Different types of biases can cause the lack of measurement invariance. There are three broad categories of biases studied in measurement invariance: construct bias, method bias and items bias.

Construct bias happens when the construct does not hold the same meaning across different cultures or there is not a complete overlap of the definition of the concept between the different groups (Byrne & Watkins, 2003, van de Vijver & Poortinga, 1997). If there is a construct bias, then there is no interpretative equivalence of the construct. Thus, no quantitative comparison can be made as it would compare two different elements (Kankaraš & Moors, 2010).

Method bias regroups the different types of biases that arise from the methodology (Byrne & Watkins, 2003). The bias can come from the sampling procedure (the different characteristics of the population are not the same in the different cultural groups) or from the administration procedure (the test is not conducted in the same condition, for example, some tests were administered in face-to-face interviews and others were self-administered online). The data could be biased either by having population with different characteristics compared or by different techniques operated to collect the data (Kankaraš & Moors, 2010).

The last category concerns items bias. This bias focuses on the different observed variables. They are considered biased if two individuals who present the same characteristics but who are from two different cultures obtain different results on the observed variables (van de Vijver & Leung, 1997). Item bias is mostly studied in item response theory. Differential item functioning (DIF) or item bias is defined as "differences in answer probabilities for respondents with equal latent disposition" (Kankaraš & Moors, 2010, p.123). Items bias does not analyse the measurement instrument as a whole but looks at the different items constituting the instrument.

2.2.4 Measurement invariance of political efficacy

Having defined political efficacy and measurement invariance, the next step is to investigate papers that have studied the measurement invariance of the political efficacy concept. In the literature, there are to our best knowledge only five articles about this topic. The first author to study measurement invariance was Mokken (1969). In his article he does not research strict invariance but only a robust scale of political efficacy which he defined as "a scale (or a factor structure) is robust for a set of cultures or nations, when its structure is approximately the same in the cultures or nations concerned." (page 426). This research does not differentiate between internal and external efficacy and the scale of political efficacy was composed of 5 questions. The robust scale is investigated in two countries, USA and Germany. The results showed that the scale was not robust.

The second research was made by LeDuc (1976). His work was in the continuity of Mokken's research. The author started from the questions developed by Mokken to create his political efficacy scale. LeDuc (1976) changed the few problematic items discovered in Mokken's study. His research was carried out only in Canada but he differentiated English speaking Canadians from French speaking Canadians.

LeDuc concluded that the scale was not robust as it did not perform well for the French Canadians compared to the English speakers. He encouraged the research of new items that could work better in cross-cultural research.

The next 3 studies took into account the fact that political efficacy was composed of two dimensions. Vecchione et al., (2014) studied only internal political efficacy. They used four items to create the latent trait, internal efficacy. Three Mediterranean countries were investigated in their study: Italy, Spain and Greece. The authors found that there was partial scalar equivalence between the three countries.

The fourth study was done by Xena (2015). She studied internal and external efficacy in the first round of the European Social Survey. There was in total 5 items defining political efficacy (3 for internal and 2 for external political efficacy) and measurement invariance was checked for 21 European countries. When the model was tested with all the countries, scalar invariance was not achieved and neither was partial scalar invariance. However, when she then tested subgroups of countries, in some cases, partial scalar invariance could be reached.

Finally, the last research on cross-cultural equivalence of the concept of political efficacy was done by Scotto et al. (2021). This time the measurement invariance is tested between two English speaking countries: Great-Britain and the United States. They used 6 items for the internal political efficacy latent variable and 4 items for the external political efficacy latent variable. In this study, the authors demonstrated that their revised indicators (in contrast to those used by Xena (2015)) give measurement equivalence.

In conclusion, the study of measurement invariance of the concept of political efficacy has been lacking. First and foremost, not a single paper used the same indicators which seems to indicate a lack of agreement on the items that measure political efficacy. Furthermore, some designs found equivalence but usually only when a small number of countries has been tested together and that those countries must have some similarities (in terms of geographical proximity, language or common history, ...).

2.3 The European Social Survey

This section describes the European Social Survey and then focuses on how this survey measures the two dimensions of political efficacy.

2.3.1 The European Social Survey

The European Social Survey is a repeated cross-sectional survey organised throughout different European countries. This survey is interested in collecting data about different topics of social sciences such as society, politics and economy. The project was initiated in 1995 and established in 2001 (*History*| European Social Survey, n.d.). The starting point of the project came from the observation that in Europe, a lack of cross-national data about socio-economic topics existed. This led to the creation of the ESS with the objective in mind to bridge this gap in European database. The project was developed with two main guidelines, first of all, the data gathered should permit high quality cross-national comparisons and secondly the data should allow the study of the evolution of European attitudes over time (ESS, 1999). Thus, the objective was to develop a survey that would collect regularly data throughout European countries at a specific time.

Since 2001, the ESS has been collecting data about European citizens' attitudes towards different core topics (such as society, politics and economy...) as well as two rotating modules. The survey is organised every two years in Europe. In practice, not all European countries take part in the ESS. However, since the establishment of the ESS, the number of participating countries has grown[1].

As one of the goals is to do cross-national comparisons, the ESS attaches a great deal of importance to the methodology (ESS, 1999). The different methodological steps are documented for every country and accessible on the website. To gather high quality data, one important methodological point was to define the eligible population and the selection procedure. For an individual to be eligible (ESS, 1999), he needs to satisfy the following criteria:

- to be a resident of one of the countries taking part in the survey
- to be at least 15 years old

This means the selection procedure includes individuals that are resident of a participating country even if they do not have the nationality or citizenship. Furthermore, there is no upper age limit in the eligibility criteria. In terms of sampling procedure, random sampling methods are used to select participants.

Before the Covid crisis, the survey was organised as face-to-face interviews (computed assisted CAPI). However, to adapt to restrictions imposed by the health crisis, in some countries the data for the tenth round was gathered by self-administrated questionnaire (either on paper or web) (ESS ERIC, 2023h).

[1]Israel is part of the ESS even if it is not in Europe

Being rigorous in the methodology is not enough to produce good data for comparison. The data quality needs to be assessed once the data has been collected. The ESS Core Scientific Team (CST) conducts data assessments on the survey (*Data Quality Assessment*| European Social Survey, n.d.). They check the measurement quality of individual questions (study the relationship between the answer to the survey and the concept), measurement invariance as well as the assessment of the socio-economic composition of the samples.

In terms of measurement invariance, the only documentation available is the report of analysis produced on some concepts from the round 1 to 5 (*ESS Rounds 1-5: Concepts, indicators, indices and their quality*| European Social Survey, n.d.). Only 5 concepts were tested: Total time spent at the media, Interest in Political issues presented in the media, Political Trust, Political Satisfaction and Quality of State Services. It appears that some of those concepts did not achieve scalar invariance. The political efficacy concept was not tested (or if it was, the information has not been made available).

2.3.2 Political efficacy in the ESS

The ESS questionnaire has included questions to measure political efficacy in almost every round since the first one in its core module "Politics". The items included in each round have evolved from one round to another. In the first round, there were 5 items related to political efficacy (ESS ERIC, 2002), 3 related to internal political efficacy and 2 to external political efficacy. In the rounds 2 to 4 (ESS ERIC, 2021, 2018a, 2008), only two questions linked to the concept of political efficacy were kept and both items were related to internal political efficacy. Regarding the rounds 5 and 6 (ESS ERIC2010; 2012), the two last items about political efficacy were excluded after an analysis showed that the two items did not measure correctly internal political efficacy (*Items dropped from the core questionnaire in ESS Rounds 5 and 6*| European Social Survey, n.d.). For the 7th round, a new set of items were introduced to study political efficacy (ESS ERIC, 2018b). This time 6 questions were added to the questionnaire to measure internal and external political efficacy (3 questions for each dimension). Building up on the round 7, four out of the 6 new items were kept for the rounds 8, 9 and 10 (ESS ERIC, 2016; 2018c; 2022). However, the scale of those four items was changed.

Finally, a number of researchers use the ESS data for their research. A few scientists have used the political efficacy items of the ESS to study political efficacy and compare results across different countries. For example: Shore et al. (2019) studied the relationship between the welfare state and level of political efficacy in different European contexts; Sulitzeanu-Kenan & Halperin (2013) research how political efficacy could be a mediator between political ideology and political preferences; Fraile & de Miguel Moyer (2022) were interested in learning more about the gender gap in internal political efficacy in Europe, etc. In none of those examples, a verification has been made about the measurement invariance of the concepts they studied.

In conclusion, the ESS is a well-known cross-national survey with high methodological aspirations. However, in practice, measurement invariance analysis is not applied to all latent concept measured in their questionnaire (or at least, results of such analysis are not accessible). A large body of scientists uses this data to produce cross-national comparisons. However, in most articles, researchers work with the assumption that it is valid to compare the countries and do not provide any analysis to verify this assumption. If measurement invariance does not hold, it is possible that results of such studies are biased. This work has the ambition to participate in the literature on political efficacy by addressing the question of the measurement invariance of the concept of political efficacy in the ESS.

2.4 Research question

The state of the art has highlighted the importance of the concept of political efficacy for the research field of political sciences. It has also shown that cross-national comparisons are frequently made in social sciences, especially in Europe. Furthermore, a well-known database to conduct such comparisons is the ESS. Political efficacy's items are part of a core module of the survey since the first round of the ESS. However, in the documentation made available by the organisation, no analysis has been produced on the measurement invariance of this concept. Only Xena (2015) proposes an analysis of this kind. Her results show that the scalar invariance is not reached for all the countries. Since Xena (2015) analysis of measurement invariance of political efficacy of the ESS, the questionnaire has evolved. However, with the new scale of political efficacy, no new analysis has been produced. A gap exists in the literature about the equivalence of the political efficacy latent construct measured by the ESS in different European countries.

Starting with the observations of this gap, it seems necessary to address the measurement invariance of political efficacy in the ESS for different reasons. First of all, as the ESS is a survey that is repeated every two years, results of analyses indicating either invariance or non-invariance are useful for the development of the questionnaire of future rounds. Secondly, most researchers do not check the assumption of measurement equivalence before making cross-national comparisons studies on political efficacy with the ESS. It is important to check the construct equivalence to remove any doubt about the validity of the conclusions reached by scientists. Finally, for future studies about political efficacy, when researchers have to decide on which data to use, knowing that the ESS produces data about political efficacy that is cross-culturally equivalent is an important characteristic to be sure that their results cannot be biased because of the instrument used to collect the data. This reflection has led us to the research question treated in this paper:

Which level of measurement invariance is reached by the concept of political efficacy in the European Social Survey in the rounds 8, 9 and 10 ?

The objective is to study the latent variables internal and external political efficacy and see if the items used in the survey are equivalent in the different European contexts. The idea is to know if the possible differences observed between countries can directly be compared or if there is a bias in the results that should be taken into account.

Chapter 3

Data

Following the description of the goal of this research, this section will present the data used in this research. The database chosen is the ESS and a selection of variables from its questionnaire has been made to answer our research question. This study will limit the analysis to the last three rounds: 8 , 9 and 10 (ESS ERIC, 2023a; 2023b; 2023c).This choice is motivated by the fact that those rounds have the same political efficacy scale in their questionnaire. As the focus of our work is political efficacy, only the items creating this scale will be looked at. In those three rounds, the ESS organisation recognised four items as being of interest for studying political efficacy in the questionnaire (*ESS Round 8 Question Design Template -New Core Items Political Efficacy*, n.d.): actrolga, cptppola, psppsgva, psppipla.

The first two items refer to internal political efficacy. The first variable, actrolga, corresponds to the question "How able do you think you are to take an active role in a group involved with political issues?". The response was in the form of a Likert scale that went from "1-Not at all able" to "5-Completely able". The second item cptppopla is "And how confident are you in your own ability to participate in politics?". The answers' scale went from "1-Not at all confident" to "5-Completely confident".

The last two items are questions related to external political efficacy. The third items (psppsgva) stands for "How much would you say the political system in [country] allows people like you to have a say in what the government does?". Once more, the possible responses were a 5 category Likert scale going from "1-Not at all" to "5-A great deal". The fourth item psppipla corresponds to "And how much would you say that the political system in [country] allows people like you to have an influence on politics?" with the same options of responses than for psppsgva.

In this paper, the names of the variables have been changed from the ones in the ESS for clarity reasons, as the ESS code name is quite long and harder to remember.

It has been decided to start by a prefix to first mark the difference between internal and external variables, the internal variables will start with an I and the external by an X. The variable actrolga has been renamed irol, cptppola became iconf, the variable psppsgva will be known as xsay and finally, psppipla has been named as xinf.

Some information about the data collected for each round is presented now. Starting with the round 8, the data was collected between august 2016 and December 2017 through face-to-face interviews. In total, the survey was conducted in 23 countries. The list of participating countries can be found in appendix A (table A1). The target response rate was 75%. In practice, this rate was not achieved and differed from one country to another. The response rate varied from 30% in Germany up to 74% in Israel (see appendix A, table A2).

Regarding the round 9, the data was likewise collected through face-to-face interviews. The time lapse during which the data was gathered started in august 2018 and finished in January 2020. There were 30 countries that took part in the ninth round of the ESS (country listed in appendix A, table A1). The target response rate was 70%. The actual response rate fluctuated between 37 and 69% in the different countries (see appendix A, table A2).

Finally, for the round 10, 31 countries participated in the tenth round of the ESS (see appendix A, table A1). Some deviation from the usual collection method happened due to the pandemic in this round.Among those countries, 22 used face-to-face interviews as the data collection method (a backup plan with online face-to-face interview was put into place) and 9 countries use self-completion questionnaire (either paper or online). The data collection period went from September 2020 to August 2022. The response rate target was 70%. In practice, the response rate ranges from 20% to 72% (see appendix A, table A2).

Chapter 4

Methodology

Having defined our research objectives, it is now time to describe which methods will be used to answer our research question. The methodology chapter is divided into three main sections. The first one is dedicated to the principal component analysis (PCA). This method is considered as a type of factorial analysis and will be used as a descriptive tool. As our analysis uses more than two variables, it becomes difficult to represent an observation when taking into account all the different variables as a graph in more than two dimensions must be constructed. The PCA permits to represent the data in a fewer number of dimensions. The idea is to visually represent the countries in a smaller dimensional space than if all the variables would be considered. The second section will describe clustering analysis. This method searches to group observations (in this case, countries) which are the most similar based on a criteria. The clustering analysis is a second descriptive tool used to represent distance (close similarity or large dissimilarity) between countries based on the different variables. Once the descriptive analysis is done, a model approach will be undertaken in the last section. The method chosen is confirmatory factorial analysis which is a model-based factor analysis. The idea is to model the relationship between our observed variables and latent constructs. The modelling of those relationships will then permit by different steps to study the measurement invariance of the scale.

4.1 Principal component analysis

4.1.1 Method

The principal component analysis is a method that can be used to visualise data without having to specify any statistical assumptions (Lebart et al. 2000). The aim of the PCA is to summarise a large dataset containing many variables into a few dimensions. The PCA permits a visualisation of variables and observations even when there is a large number of variables. Furthermore, the analysis can identify patterns in the data such as variables that are linked or that measure the same phenomenon as well as grouping observations who have similar values in the different variables (Husson et al. 2017).

The idea behind the PCA is to find an axis that is a linear combination of the different variables to summarise in the best way the cloud of observations in the p-dimensional space (Husson et al. 2017). The first component should explain the maximum variance of the data possible. Then, the next component is computed to find the best new dimension that is orthogonal to the first one which would explain the most variance of the data and so on. However, there is a hierarchy in the components such that the first component will explain more variance than the second and the third will explain less variance than the second but more than the fourth, etc. In PCA, the number of principal components computed is equal to the number of variables. Usually, only the first few principal components are used in order to visualise the data as they explained the most variance.

Principal component analysis consists of applying (in general) factorial analysis on centered and standardised data matrix (nxp, n= number of observations and p= number of variables):

$$Z = \frac{1}{\sqrt{n}}\mathbf{X}_{cs} = \left(\frac{x_{ij} - \overline{x}_j}{\sqrt{n}s_j}\right)_{\substack{i=1,\ldots,n \\ j=1,\ldots,p}} \tag{4.1}$$

Where

- n is the sample size
- $\mathbf{X}_{cs}$ is the centered and standardized matrix of observations
- x_{ij} is the i_{th} observation of the j_{th} variable
- $\overline{x}_j$ is the sample mean of the j_{th} variable
- s_j is the standard deviation of the j_{th} variable

To compute principal component analysis (Husson et al. 2017), it is necessary to compute the eigenvalues and eigenvectors of the sample correlation matrix (by diagonalisation) which is computed as:

$$Z'Z = R \tag{4.2}$$

Where

- Z is the centered and standardized matrix
- R is the sample correlation matrix

Having obtained the eigenvalues, they are then ranked in decreasing order. The eigenvalue is interpreted as the “explained variance” of the component. Thus, the first component is computed based on the eigenvectors corresponding to the eigenvalue with the highest value. The Principal component is the new variable that corresponds to the multiplication of the scaled and centered data by the eigenvectors.

The principal component is

$$y_{\alpha i} = \sum_{j=1}^{p} \frac{x_{ij} - \overline{x}_j}{s_j} u_{\alpha j} \tag{4.3}$$

$$\mathbf{y}_\alpha = \mathbf{X}_{cs}\mathbf{u}_\alpha : n \times 1 \tag{4.4}$$

Where

- $y_{\alpha i}$ is the coordinate of the i_{th} observation on the α principal component
- $u_{\alpha j}$ is the j^{th} eigenvector of R

The PCA permits a representation of the variables and observations. Firstly, for the observation’s representation, the graph represents the components wanted, and the observations are represented with their coordinates of the components. To read this graph, two points will be close if they have similar values for the different variables that contribute to the component analysed.

Secondly, the PCA offers a representation of the variables. It is done through a correlation circle. Again, the two axes represented on the graph are the two principal components. Then, the variables are represented with coordinates on each axis which are the correlation between the variables and the principal component. Two variables will be close in the graph, if for the majority of the individuals, they obtain similar values on both variables.

4.1.2 Implementation

Before applying the PCA, only the individuals that respond to the four questions about political efficacy were selected. Having selected the individuals, then aggregation of the results was computed[1]. Thus, for every country participating in the ESS, for each question, five new variables were created, one for each possible answer. These new variables consist of the percentage of the population of the country to have answered to that particular category.

Once this pre-treatment of the data has been done, a PCA has been computed using the function *PCA* of the *FactoMineR* package. The PCA has been fitted on all three rounds of data. However, countries that did not participate in all three rounds have been added as supplementary observations in the PCA procedure. The variable plot as well as the individual plot have been produced and analysed (see results chapter).

4.2 Clustering

4.2.1 Method

Clustering is a data analysis method that creates partition of the data to regroup observations that are similar. As for the PCA, the clustering method does not have any assumptions made on the data before implementing the method. There are different types of clustering. In this thesis, hierarchical clustering will be used. It is a type of clustering that is exploratory in the sense that it starts with the same number of clusters as the number of observations and finishes with only one cluster containing all the observations (Lebart et al. 2000).

Clustering methods are based on a function of dissimilarity between observations. However, how to compute the dissimilarity between those ? The most common measure for dissimilarity is to measure the distance between points. There are different types of metrics for distance. In this work, the Euclidean distance will be used because it is the distance used in the Ward's criterion (see below).

Having defined how the distance is computed, the next step is to decide which rule for agglomeration of the cluster will be used, the dissimilarity function. This means what are the criteria used to select which clusters to merge. There exist different rules, such as simple linkage, complete linkage, etc. In this paper, the Ward's minimum variance criterion will be used as it is the only one which is based on

[1]The aggregation has taken into account the survey design to compute the proportion. The *survey* package was used to do it.

variance between clusters. This criterion is based on the concept of inertia which is defined as the sum of the variances (Husson et al. 2017).

$$\begin{aligned}\mathcal{I}_T &= \frac{1}{n}\sum_{i=1}^{n}\|x_i - \overline{x}\|^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p}(x_{ij} - \overline{x}_j)^2 = \sum_{j=1}^{p} s_{ij}\end{aligned} \tag{4.5}$$

Where

- n is the sample size
- x_{ij} is the i^{th} observation of the j^{th} variable
- $\overline{x_j}$ is the mean of the j^{th} variable
- s is the variance

The inertia is then decomposed in two elements, the first one is the within inertia which consists of measuring the variance between the points of a same cluster:

$$\mathcal{I}_W = \sum_{k=1}^{q}\frac{n_k}{n}\frac{1}{n_k}\sum_{i\in I_k}\|x_i - \overline{x}_k\|^2 = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - \overline{\mathbf{x}}_{k(i)}\|^2 \tag{4.6}$$

Where

- n_k is the sample size of group k
- $\overline{x}_k$ is the mean of group k

The second element is the between inertia which measures the variance between the different clusters:

$$\mathcal{I}_B = \sum_{k=1}^{q}\frac{n_k}{n}\|\overline{x}_k - \overline{x}\|^2 = \frac{1}{n}\sum_{i=1}^{n}\|\overline{\mathbf{x}}_{k(i)} - \overline{\mathbf{x}}\|^2 \tag{4.7}$$

The Ward criterion consists of finding the aggregation of clusters that minimises the within inertia and maximises the between inertia (Husson et al. 2017). When starting a hierarchical cluster, each observation is a cluster, the within inertia is equal to 0 and the between inertia is the total variance whereas at the end of the algorithm, the within inertia is equal to the total variance and the between inertia is equal to 0. As each time, you aggregate a new point in a cluster, the within inertia will go up and the between inertia will go down. The objective of this criterion is to find the merger of two clusters that minimises the transfer of the dissimilarity function from between to within inertia.

The hierarchical clustering is computed by an algorithm that follows the steps presented below. Before computing the different steps of the algorithm (Lebart et al. 2000), the choice of which dissimilarity function to use need to be made.

1. The first step is to start with the partition where each observation corresponds to a cluster.
2. The second step is to compute the proximity matrix
3. Find the minimal value of the proximity matrix and aggregate the two clusters who obtained this value.
4. Next compute again the proximity matrix, by deleting the rows and columns of the two clusters merged and computing the row and column for the new cluster.
5. Finally, repeat the step 3 and 4 until there is one cluster containing all the observations.

4.2.2 Implementation

The same pre-treatment than in the PCA has been applied on the data before applying the cluster analysis. One analysis has been made for the three rounds of the ESS and only the common countries in the three were taken into account. The clustering has been computed using the function *hclust* from the *stats* package and using the *ward* criterion. The result has been plotted in dendrogram (see results chapter).

4.3 Confirmatory factor analysis

4.3.1 Structural Equation Modelling

Structural Equation Modelling (SEM) is a family of methods that is interested in modelling relationships between observed and latent variables. It is a methodology that is driven by theoretical assumptions as SEM are explicit models meaning that the researcher defines all the parameters of the models. The idea behind such models is to be able to "test a theory by specifying a model that represents predictions of that theory among plausible constructs measured with appropriate observed variables" (Kline, 2016, p.10).

SEM models are models that differentiate between observed variables and latent variables (Kline, 2016). The observed variables are the variables that the researchers can directly measure. A latent variable is an hypothetical construct that is usually an abstract concept that cannot be directly observed by a researcher. Consequently, to measure a latent variable, a group of observable variables that capture different aspects of the latent construct is constituted. One major difference between observed and latent variable in SEM models is that latent variables can only be continuous variables whereas observed variables can either be continuous or categorical variables.

SEM models are separated between two different categories of models: measurement models and structural models (Brown, 2015). Measurement models are models that study the relationship between observed variables and latent variables. In contrast, structural models focus on the relationships between latent variables. Regarding the research objective of this work, measurement models are the ones that will be used.

The specification of the causal relationships sometimes leads to a misunderstanding of what SEM models actually can do (Kline, 2016). If the model specified does fit the data, it does not prove that the researchers' model is the true model explaining the phenomenon. There could be equivalent models that explain just as well the data. However, the SEM model can be used as a disconfirmatory model in the sense that if the model specified does not fit the data, this helps us reject a false model. Bolen (1989) explained this as: "If a model is consistent with reality, then the data should be consistent with the model. But if the data are consistent with the model, this does not imply that the model corresponds to reality" (page 68).

4.3.2 Confirmatory factor analysis

Confirmatory factor analysis (CFA) is a measurement model of the SEM class. In measurement models terminology, observed variables are called indicators and latent variables are called factors. The latent variables are understood as hypothetical constructs that explained the correlation between multiple indicators. The factor would be a variable that cannot be observed but which would influence different observed indicators (Brown & Moore 2012).

Before diving into the CFA model, the classical measurement theory permits a first approach to conceptualise the relationship between indicators and factors (Roos & Bauldry, 2022). In this framework, an indicator can be divided into two elements. The first component would be the true component, understood as the part of the observed variable that captures the information about the factor. The second element is the error component, meaning that it expresses a source of variation in the indicator that is not linked to the factor. This framework is an introduction to the methodology used in CFA as it only explains the relationship between one indicator to one factor.

In CFA methodology, the relationships studied are the ones between multiple indicators and one or multiple factors. The CFA method builds on the common factor model which believes that indicators can be represented as a linear combination of common factors (Kline, 2016). In this framework, the variance of each indicator is separated in two elements: the common variance and unique variance. The common variance corresponds to the variation in the observed measure that can be attributed to a factor which is common to a few other observed measures. The unique variance represents the variance that cannot be explained by the common factor. This unique variance is composed of a combination of the random error variance as well as the variance that is specific to each indicator. In a CFA model, the relationships between indicator and factor are specified a priori.

Based on the common factor method, a standard CFA model can be written as a set of linear equations (Widaman & Reise, 1997). Those equations describe the relationship between the indicator and the factor where the factor is considered as the explanatory variable of the indicator. Thus, this can be written as:

$$y_{ij} = \lambda_{i1}\eta_{ij1} + \lambda_{i2}\eta_{ij2} + \cdots + \lambda_{im}\eta_{ijm} + \varepsilon_{ij} \tag{4.8}$$

Where

- y_{ij} represents the j^{th} observation of the i^{th} indicator
- λ_{im} represents the coefficient slope (also known as the loading) of the m^{th} factors on the i^{th} indicator.
- η_{ijm} represents the value of the m^{th} factor for the j^{th} observation of the i^{th} indicator
- ε_{ij} represents the error of the model for the j^{th} observation of the i^{th} indicator

A more compact form of this model is:

$$Y = \Lambda\eta + \varepsilon \tag{4.9}$$

- Y is (p x 1) matrix of the centered values on the p observed variable for the i^{th} individual

- Λ is the (p x m) loadings matrix of the p observed variables and the m latent indicators.

- η is (m x 1) matrix of score on the m latent indicator of the i^{th} observation

- ε is the (p x 1) column vector of the error terms of the i^{th} observation

From this equation, it is possible to obtain a covariance structure model by multiplying each side of the equation by its transpose and then dividing both sides by (N-1), this leads to the next equation:

$$\Sigma = \Lambda \Psi \Lambda' + \Theta \tag{4.10}$$

- Σ is the covariance matrix of the indicators,

- Λ is the coefficient matrix,

- Ψ is the covariance matrix between the factors

- Θ is the covariance of the error terms.

The linear equation previously can be modified to include the intercept (and latent mean) of the model (Widaman & Reise, 1997). In the previous equation, the intercept was not included as the y was actually the centered observation of the indicator. If the observation is not centered, the equation becomes:

$$Y = \tau + \Lambda \eta + \Theta \tag{4.11}$$

$$M_Y = \tau + \lambda + \kappa \tag{4.12}$$

- τ is the intercept

- M_Y is the indicator mean

- κ is the mean of the latent variable (factor)

4.3.3 Specification of the model

The first step to compute a CFA model is to specify the model (Kline, 2016; Roos & Bauldry, 2022). The researcher has to have thought about a number of elements to specify the model:

- Is there one or more latent variables?
- Which indicators measure which factors?
- Are the error of the indicators independent from one another or is there for some variables correlated errors ?
- If there is more than one factor, what is the relationship between the factors ?
- Etc.

4.3.4 Identification of a model

Having defined what a CFA model is, the next step is to identify the model. The identification of a model is necessary to obtain an unique estimation of the different unknown parameters of the model. In CFA, a specific issue concerning model identification exists, the scaling of the latent variable. As the metric of the unobservable variable is unknown, the scientist has to define the measurement metric of the factor. In CFA, there are two different methods to do this(Kline, 2016; Roos & Bauldry, 2022; Brown, 2015):

- The first one is the reference or marker variable. This method consists in assigning the factor loadings of one indicator to 1 and its intercept to 0. This method gives to the latent variable the same metric as the observed variable that has been chosen as reference. This method gives unstandardised and standardised solutions of the parameters.
- The second method consists of fixing the variance of the factor to 1 and its mean to 0. However, this method only produces standardised solutions.

Secondly, there is also the question of the statistical identification of the model. Statistical identification refers to the ability of the model to identify a unique solution to the parameter estimation. It is based on the difference between the number of free parameters that have to be estimated and the number of known information available (in the covariance matrix and the means of the indicator). To check the statistical identification of a CFA model, the degrees of freedom of the model should be looked at (Kline, 2016; Roos & Bauldry, 2022; Brown, 2015).

The degree of freedom is defined as the number of known parameters minus the number of unknown parameters. If the number of degrees of freedom is equal to 0, then the model is just-identified. If the degrees of freedom are a positive number, then the model is over-identified. And if the number of degrees of freedom of the model is less than 0 then the model is under-identified. An under-identified model does not have a solution because different parameter estimations could lead to the same model fit. Thus, only just-identified and over-identified models obtain a solution. Based on this statistical identification, a rule of thumb for the minimum number of indicators by factor has been created (Kline, 2016). For a one factor model, there needs to be at least three indicators in the model. For a multiple factor model, each factor needs to be related to at least two indicators.

However, in case of a theoretically identified model, the issue of empirical under-identification still exists. An empirical under-identified model is defined as "a measurement model is identified but specific parameter estimates from a given source of data lead to a not identified model" (Roos & Bauldry, 2022, p.34). With this risk in mind, in the case of a multiple factor model, it is recommended that each factor should be related to at least three indicators.

4.3.5 Estimation of the model

Once specification and identification steps of the model are completed, the model needs to estimate the different parameters. The specification step has introduced in the model various relationships between indicators and factors. This information is summarised in the implied covariance matrix of the model (and implied mean vector). The estimation procedure of the CFA method uses this information. To estimate the parameters, the CFA procedure researches the value of the parameters which reproduce the best the sample covariance matrix (Roos & Bauldry, 2022).

$$S \simeq \hat{\Lambda}\hat{\Psi}\hat{\Lambda}' + \hat{\Theta} = \hat{\Sigma} \tag{4.13}$$

- S is the sample matrix covariance
- $\hat{\Sigma}$ is the estimated covariance matrix of the indicators,
- $\hat{\Lambda}$ is the estimated coefficient matrix,
- $\hat{\Psi}$ is the estimated covariance matrix between the factors

- $\hat{\Theta}$ is the estimated covariance of the error terms.

The estimation of the parameter of a CFA model requires a fitting function (Lei & Wu, 2012). The idea is to minimise some form of discrepancy between the sample variance-covariance matrix and the implied variance-covariance matrix of the model. Following the choice of an estimator, the fitting function will be different. The estimation in CFA searches to find value for the different unknown parameters that will lead the estimated implied covariance matrix to be as similar as possible to the sample covariance matrix. In CFA, it is not always possible to obtain a closed-form solution and thus most estimation methods in CFA rely on an iterative procedure (when a model is over-identified, an iterative procedure is necessary).

The most common estimator in CFA is the maximum likelihood estimator (Lei & Wu, 2012, Chou & Bentler 1995), which is:

$$F_{ML} = ln|S| - ln|\Sigma| + trace[(S)(\Sigma^{-1})] - p \tag{4.14}$$

- $|S|$ is the determinant of the matrix S
- $|\Sigma|$is the determinant of the Σ matrix
- p is the number of indicators of the model

The maximum likelihood is widely used in the case where the indicators are normally distributed continuous variables because of its properties (Lei & Wu, 2012, Roos & Bauldry, 2022). Firstly, maximum likelihood estimator computes standard errors for the parameters estimated. Secondly, maximum likelihood permits to test the fit of the model because it has a Chi square distribution when you multiply the fitting function by $N - 1$. Finally, it is an estimator that is asymptotically unbiased, consistent and asymptotically efficient.

However, the maximum likelihood estimator is based on the assumption that the observed variables follow a multivariate normal distribution (Roos & Bauldry, 2022). When using maximum likelihood with non-normal variables, there is a risk of obtaining biased standard errors and model fit or even incorrect parameters estimates (Li, 2016; Finney & Disteffano 2006). Finney & Disteffano 2006 have summarised results about using maximum likelihood estimator on non-normal data (continuous or categorical). It appears that when the ML is used on non-normal continuous data, the fit of the model and the standard error are biased but no-bias is detected on the parameters estimates. When using ML with categorical ordinal data, the model fit, standard error as well as parameter estimations are biased. When dealing with non-normally distributed data, the two most used alternatives are either using a robust maximum likelihood estimator or using an estimator that does not make any distribution assumptions about the data.

Robust maximum likelihood estimates the parameter just like the maximum likelihood estimator. The difference is that the chi square statistic and the standard error are statistically corrected to adjust for the non-normality of the data. A well-known correction is the Satorra-Bentler correction. The corrected standard errors are computed using the robust covariance matrix:

$$\left(\check{\Delta}'\check{W}^{-1}\check{\Delta}\right)^{-1}\check{\Delta}'\check{W}^{-1}\check{\Gamma}\check{W}^{-1}\check{\Delta}\left(\check{\Delta}'\check{W}^{-1}\check{\Delta}\right)^{-1} \tag{4.15}$$

- $\check{\Gamma}$ is the estimated asymptotic covariance matrix of S
- $\check{\Delta}$ is the matrix of model derivatives evaluated at the parameter estimates
- $\check{W}$ is the estimated weight matrix

The Satorra-Bentler chi square statistic (also called mean corrected (scaled) chi-square) formula is:

$$T_{ML-M} = \frac{df}{tr(\check{U}\check{\Gamma})}T_{ML}, \qquad where \qquad \check{U} = \check{W}^{-1} - \check{W}^{-1}\check{\Gamma}\left(\check{\Delta}'\check{W}^{-1}\check{\Delta}\right)^{-1}\check{\Delta}\check{W}^{-1} \tag{4.16}$$

With df as degrees of freedom. This method is recommended when the observed indicators are continuous variables which are not normally distributed.

The estimator with a non-distributional assumption used CFA is the weighted least square estimator (Finney & Disteffano 2006). The WLS estimator formula is:

$$F_{WLS} = [S - \Sigma]'W^{-1}[S - \Sigma] \tag{4.17}$$

- W^{-1} corresponds to the inverse weight matrix that contains the asymptotic covariance matrix
- S is the matrix of the sample covariance matrix
- Σ corresponds to the model implied covariance matrix.

Muthén (1984) has developed from the WLS approach an estimation method that takes into account the categorical characteristics of ordinal data. The idea is to consider that the categories of an indicator represent regions on continuous scale, demarcated by thresholds. This continuous scale would follow a normal distribution. For example, with the variable “how confident are you in your own ability to participate in politics?”, the researchers would believe that there exists a continuum from low confidence to high confidence and that the categories represent different regions of this continuum (Roos & Bauldry, 2022).

In practice, this method considers that for each indicator, there exists a latent response variable (can be viewed as a latent indicator) that is a continuous normally distributed variable that describes which amount of this latent indicator is necessary to answer a certain category on the observed indicator (Kline, 2016). Thus, a threshold is computed for every n-1 category of the ordinal indicator. This conception leads the CFA model to study the linear relationship between the latent indicator and the factor (Roos & Bauldry, 2022). The linear equation of a categorical CFA are:

$$\begin{aligned} X_j &= c, \quad \text{if } \tau_{j,c} < X_j^* < \tau_{j,c+1} \quad c = 0, 1, \ldots, C-1 \\ X_j^* &= \lambda_j \eta + \varepsilon_j, \quad j = 1, 2, \ldots, J \end{aligned} \tag{4.18}$$

- X_j is the j_{th} categorical observed variable
- C is the category
- τ is the threshold
- X_j^* is the latent response variable of the j_{th} categorical observed variable
- λ is the factor loading
- η is the latent factor
- ε is the error

The ordinal CFA also needs an additional identification (Brown, 2015). Indeed, the latent response variable X^* needs to be scaled. There are two methods to scale this new latent response variable:

- The first one is the delta parameterization which consists of specifying that the total variance of the X^* variable is equal to 1.

- The second method is the theta parameterization. This method is different as it only fixes the residual variance of the X^* variable to 1.

The WLS approach with categorical variables exchanges the sample covariance matrix by the polychoric correlation which would contain the information about the relationship between the latent indicator. The polychoric correlation uses the information contained in the different bivariate contingency tables obtained from the indicators to estimate the correlation. Olsen (1979) describes a two-stages procedure to calculate the polychoric correlation. The first step is to estimate the thresholds with the cumulative response distribution (Flora & Curran, 2004):

$$a_i = \Phi_1^{-1}(P_{i.}) \quad \text{and } b_j = \Phi_1^{-1}(P_{.j}) \tag{4.19}$$

- a_i is the threshold i of the variable $X1$
- b_j is the threshold j of the variable $X2$
- P_{ij} is the observed proportion in cell (i, j) of the bivariate contingency table
- $P_{i.}$ and $P_{.j}$ are the observed cumulative marginal proportions
- Φ_1^{-1} is the inverse of the normal cumulative distribution function

The second step uses both the thresholds estimates and the bivariate contingency table to estimate the correlation between the two latent indicators (Flora & Curran, 2004). The estimation is done via maximum likelihood. The log-likelihood of the bivariate sample formula is:

$$l = \ln\, K + \sum_{i=1}^{s}\sum_{j=1}^{r} n_{i,j} \ln\, \pi_{i,j} \tag{4.20}$$

- K is a constant
- $n_{i,j}$ is the observed frequency of the cell (i, j) of the contingency table
- $\pi_{i,j}$ is the probability that a given observation would be in cell (i, j)

Olsen has given the formula for the π:

$$\pi_{i,j} = \Phi_2(a_i, b_j) - \Phi_2(a_{i-1}, b_j) - \Phi_2(a_i, b_{j-1}) + \Phi_2(a_{i-1}, b_{j-1}) \tag{4.21}$$

- Φ_2 is the bivariate normal cumulative density function with correlation ρ.

The maximum likelihood of the ρ parameters is the polychoric correlation between the variable $X1$ and $X2$. The WLS estimator with categorical data is:

$$F_{WLS} = [s - \sigma(\theta)]' W^{-1} [s - \sigma(\theta)] \tag{4.22}$$

- W^{-1} is the matrix containing the sample statistics, meaning the sample estimation of the polychoric correlation and thresholds
- s is the matrix containing the vectorized elements of the model implied variance-covariance matrix
- $\sigma(\theta)$ is the inverse of the weight matrix, corresponding to the asymptotic covariance matrix of the polychoric correlation

In practice, the full weighted least square is rarely used because as the number of indicators increases, it needs a certain level of computer resources (as the weight matrix becomes larger). Furthermore, to obtain stable parameters estimate, the model needs large sample (Flora & Curran, 2004).

A robust weighted least square estimation, also called diagonal least square estimator has been developed to avoid problems of WLS estimator. The robust WLS estimator is based on a similar idea of the WLS estimator. However, the weight matrix is different. Instead of using the inverse of the whole asymptotic covariance matrix as the weight matrix, the robust estimator only uses the diagonal of this matrix. Thus, the robust WLS estimator is:

$$F_{WLSMV} = [s - \sigma(\theta)]' W_D^{-1} [s - \sigma(\theta)] \tag{4.23}$$

- W^{-1} is the matrix containing the sample statistics, meaning the sample estimation of the polychoric correlation and thresholds
- s is the matrix containing the vectorized elements of the model implied variance-covariance matrix
- $\sigma(\theta)$ is the inverse of the diagonal weight matrix, corresponding to the asymptotic variances of the polychoric correlation

The robust estimator then computes robust standard errors and chi-square statistics.

4.3.6 Fit of the model

Once the model has been estimated, it is possible to evaluate the fit of the model. As the objective of the CFA model is to obtain a model's implied variance-covariance matrix that is as similar as possible to the sample variance-covariance matrix, the fit of the model is evaluated based on those two covariance matrices (Widaman & Reise, 1997). If the implied covariance matrix is equal to the sample covariance matrix, then the model perfectly fits the data. There are two mains strategies to evaluate the fit of the model. The first one is to conduct a chi-square test and the second is to compute fit indices based on the chi-square statistics. Both strategies are based on the chi square statistic, so it will be first defined.

When using maximum likelihood estimator, the chi-square statistic is computed by multiplying the fitting function (evaluated in the values that minimise the function) by the sample size minus 1 (Roos & Bauldry, 2022). The degree of freedom of the chi-square distribution of the statistic is obtained by the difference between the number of distinct elements in the sample covariance matrix and the number of parameters:

$$T_{ML} = (N-1)F_{ML} \tag{4.24}$$

- T_{ML} is the statistics obtained with the maximum likelihood estimator
- F_{ML} is the value obtained when the ML fitting function is minimized
- N is the sample size

When a robust estimator either for maximum likelihood or weighted least square is used, the chi-square statistic is obtained by correcting the fitting function value for non-normality. In the literature, the test is not recommended as it has limitations based on the sample size (D'Urso et al. 2022). The key limitation is that as the sample size grows, the power of the test also grows. As the power of the test grows, the test is able to detect negligible deviation from the strict equality between the two covariance matrices (null hypothesis). This leads to the conclusion that the model does not fit the data even if the differences are minor (Roos & Bauldry, 2022). Thus, the $\mathcal{X}^2$ test will not be used in this paper.

A first type of indice is a misfit indice (D'Urso et al. 2022). This type of indice does not focus on the exact fit but analyses the deviation of the model from an exact fit. The most used misfit indice is the RMSEA (root mean square error of approximation). The RMSEA is an indice that is based on the $\mathcal{X}^2$ statistics and number of degrees of freedom.

$$RMSEA = \sqrt{\frac{\delta}{df(n-1)}} \quad \text{where } \delta = \mathcal{X}^2 - df \tag{4.25}$$

An RMSEA that is smaller than 0.05 is considered as a "good" fit, whereas a value between 0.05 and 0.08 is considered as "acceptable" fit but a value larger than 0.08 indicates a "poor" fit of the model (D'Urso et al. 2022).

A second category of indices is a comparative fit indice (D'Urso et al. 2022). Here, a comparison is made between the chosen model and a null model. One of the most used indices of this category is the comparative fit index (CFI) which is computed as:

$$CFI = \frac{\delta(\text{Baseline}) - \delta(\text{User})}{\delta(\text{Baseline})} \quad \text{where } \delta = \mathcal{X}^2 - df \tag{4.26}$$

The baseline model, also called the null model, corresponds to the model where the specification made is that there is no common factor between the observed indicators (Roos & Bauldry, 2022). A model is considered to have an "acceptable" fit if its CFI value is larger than 0.9 and a "good" fit if the CFI value is higher than 0.95 (D'Urso et al. 2022).

4.3.7 Measurement invariance in CFA

Measurement invariance can be tested with a CFA model. As explained before, there are different levels of invariance. The CFA permits to test one level after another. The logic is the same for both continuous and categorical CFA models. The idea is that by applying equality constraints on different parameters of the model, it is possible to check the different levels of measurement invariance. This will lead to nested models. To test measurement invariance in CFA, a series of models are fitted and then compared to see if the invariance can be accepted or not. In CFA methodology, to test measurement invariance a multi-group CFA model is needed. A multi-group CFA is a CFA model that fits the same model to different samples of data at the same time and permits free estimation of the parameters for each group or permits to put equality constraints on parameter for the different groups. The equations of the multi-group CFA model are:

$$Y = \Lambda^{(g)}\eta + \varepsilon \tag{4.27}$$

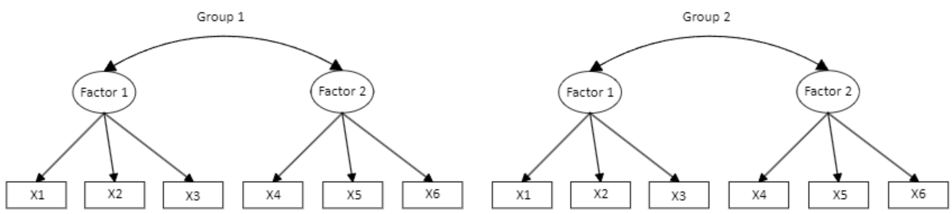
$$X_j \quad = \quad c, \quad \text{if } \tau_{j,c}^{(g)} < X_j^* < \tau_{j,c+1}^{(g)} \quad c = 0, 1, \ldots, C-1 \tag{4.28}$$

$$X_j^* \;\; = \;\; \lambda_j^{(g)}\eta + \varepsilon_j, \quad j = 1, 2, \ldots, J$$

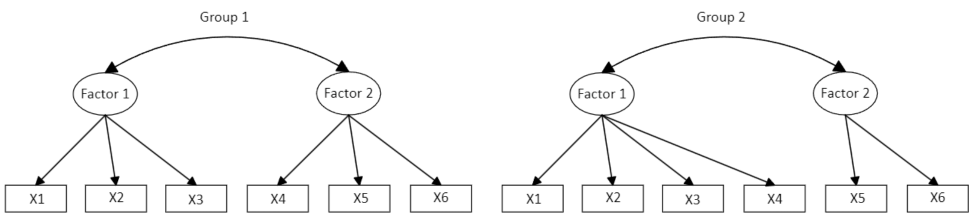
The equations are the same as 4.9 & 4.18, the only difference is the indice g that is added to show that it corresponds to the parameter from the g sample.

The different parameters of a standard multi-group CFA model (a model where each indicator can only be linked to one factor) are the factor loading, the unique variance, the factor variance, the factor covariance (when more than one factor in the model) and the intercept. The factor loading corresponds to the regression slope of the factor that explains the indicator. The unique variance corresponds to the variance of one indicator not explained by the model, also called the measurement error (as this unique variance is attributed in general to measurement error). The factor variance corresponds to the variability of the population sampled on the factor. The factor covariance is the parameter that estimates the relationship between the different latent variables. (Brown & Moore 2012). The measurement invariance for the continuous CFA model will first be presented. Once it is done, the measurement invariance steps for the categorical model will be explained.

Starting with configural equivalence, the idea is to check whether the construct is measured by the same observed variable in the different groups. Visually, configural measurement invariance checks that the factors are related to the same indicators for every group. For example, see figure 4.1:



(a) Invariance

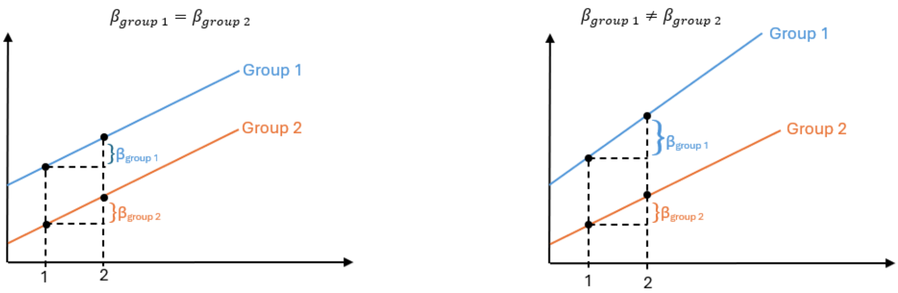


(b) Non-invariance

Figure 4.1: Configural invariance

To compute configural invariance, the same model is fitted to the different samples and the parameters are freely estimated. To judge on the configural invariance, the fit of the data needs to be looked at (by looking at different fit index previously explained). If the model fits the data, configural invariance can be accepted and the next level of invariance will be looked at.

Metric equivalence is interested in the strength of the relationship between indicators and factors. In a CFA, the strength of the relationship is computed by the loadings parameters. Visually, two groups would achieve metric invariance if the coefficient (of the regression of an indicator on the factor) is equal between the two groups. For example, see figure 4.2 :

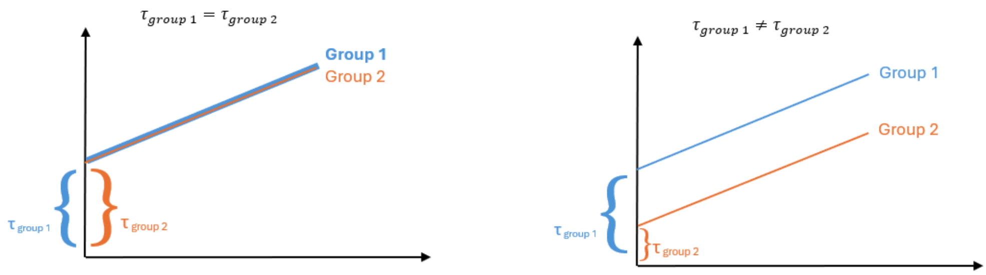


(a) Invariance (b) Non-invariance

Figure 4.2: Metric invariance

Metric invariance is tested by adding equality constraints on the unstandardised factor loading of the model. When the new model is fitted, the difference between the different indices of quality of the model will be assessed. If the model does not fit as well as the configural invariance model, the metric invariance model is rejected.

Finally, scalar invariance is interested in the origin of the scale. In CFA, the origin of the scale corresponds to the intercepts of the model. Visually, two groups would reached scalar invariance if the intercepts (of the regression of an indicator on the factor) are equal. For example, see figure 4.3 :



(a) Invariance (b) Non-invariance

Figure 4.3: Scalar invariance

The scalar invariance model is created by adding equality constraints on the intercepts on the metric invariant model. To accept the hypothesis of scalar invariance, the model fit cannot become poorer than the fit of the metric invariant model, this is decided based on the difference in terms of fit indices.

For the categorical CFA model, the configural and metric equivalence steps are the same. However, for the scalar invariance step, there is a difference. Indeed, instead of putting an equality constraint on the intercepts, the equality constraint is put on the thresholds. The intercepts are not investigated as when computing the configural model, the intercepts are set to 0 when fitting an ordinal model.

Researchers have studied differences in fit index especially to test measurement invariance and cut-off values have been established both for RMSEA and CFI. In this work, the 2 index strategy proposed by Hu & Bentler (1999) will be applied. This strategy consists in rejecting the invariance hypothesis if for the two fit indices the invariance hypothesis is rejected. If the delta RMSEA between an constrained and unconstrained model, is higher than 0.01 then the invariance is rejected for the RMSEA (Cheung & Rensvold, 2002).

$$\Delta\text{RMSEA}_{metric} = \text{RMSEA}_{metric} - \text{RMSEA}_{configural} \quad (4.29)$$

$$\Delta\text{RMSEA}_{scalar} = \text{RMSEA}_{scalar} - \text{RMSEA}_{metric} \quad (4.30)$$

If the delta CFI is larger than -0.01 than the invariance has to be rejected for the CFI (Cheung & Rensvold, 2002).

$$\Delta\text{CFI}_{metric} = \text{CFI}_{metric} - \text{CFI}_{configural} \quad (4.31)$$

$$\Delta\text{CFI}_{scalar} = \text{CFI}_{scalar} - \text{CFI}_{metric} \quad (4.32)$$

4.3.8 Invariance or non-invariance ?

Once the measurement invariance qualities of an instrument are investigated, what happens? There are two different possible scenarios. The first one would be that the instrument achieves scalar invariance. In this case, it is valid to compare latent means across the different groups. The multi-group CFA is able to compute the latent means. The procedure fixes the latent means of one group to zero and the means of the other group are freely estimated. The latent mean value of a non-reference group corresponds to the difference between the reference group and this group.

The second scenario would be that the instrument does not reach scalar invariance. In this case, most researchers look to see if partial invariance could be reached. However, when there are less than three indicators by factor, partial invariance cannot be tested. If scalar invariance cannot be reached, researchers would try to modify the instrument to get one that would be scalar invariant. However, no researcher usually questions the level of non-invariance. Indeed, as measurement invariance CFA tests strict equality, it is possible that scalar invariance would not be reached but in practice the difference would be very small. If scalar invariance is not reached in this work and as partial invariance is not possible in this case, effort to represent visually the invariance will be deployed. The idea is that the CFA model is able to compute for every observation a latent factor score. A plot would be drawn where the x axis would be the latent score and the y axis would be one of the indicators. The different groups would be plotted in different colours to see how the relationship between the latent factor and the indicator is similar or not across groups.

4.3.9 Implementation

The data used for the CFA analyses consists of the observations in 21 countries that took part of all three rounds of the ESS. Furthermore, only observations where no missing data was present in the different variables used in the model were selected. The CFA analyses were produced with the function *cfa* of the package *lavaan* in R. Firstly, the structure of the model had to be specified in an object before assigning this object as the *model* argument in the *cfa* function. For the estimation with the maximum likelihood estimator with the Satorra-Bentler correction, the *test* argument was specify with the "satorra-bentler" test. For the estimation with robust DWLS estimator, the estimator argument was used to specify the "WLSMV" estimator. Furthermore, to specify that the data is ordinal the argument *ordered* was used. When multigroup CFA was computed, the argument *group* was specified in the cfa function. To constraint the different models to test for measurement invariance, the argument *group.equal* was specified. The argument *sampling.weights* was used to specify the sampling weights. The fit measure RMSEA and CFI were computed by the model.

Chapter 5

Presentation of the Results

This section starts with a descriptive graph showing the means of the different countries and rounds. Next, the results of the principal component analysis will be described before presenting the cluster analysis results. Once the results of the different descriptive tools have been established, the results about the CFA analysis will be shown. The continuous analysis will first be presented followed by the categorical analysis results.

5.1 Descriptive graph

Figure 5.1 represents the mean of external and internal political efficacy for each country and for every round. Internal efficacy is computed as the sum between irol (take an active role in a political group) and iconf (confidence to participate) and external efficacy is computed as the sum of xsay (have a say in the government activities) and xinf (influence in politics)[1]. A different symbol has been chosen to distinguish between rounds and each colour corresponds to one country[2]. On the graph, a difference has been made between countries that are present in all three rounds and the ones which were not. The countries in all three rounds are less transparent than the other ones. Finally, the countries which are in a dot line do not actually represent countries but they represent the different Belgian regions (reason why they are in the same colour as Belgium).

[1]The mean computation has taken into account the survey design.

[2]This graph is an estimation of internal and external efficacy where each observed variable has the same weight in the computation of the means.

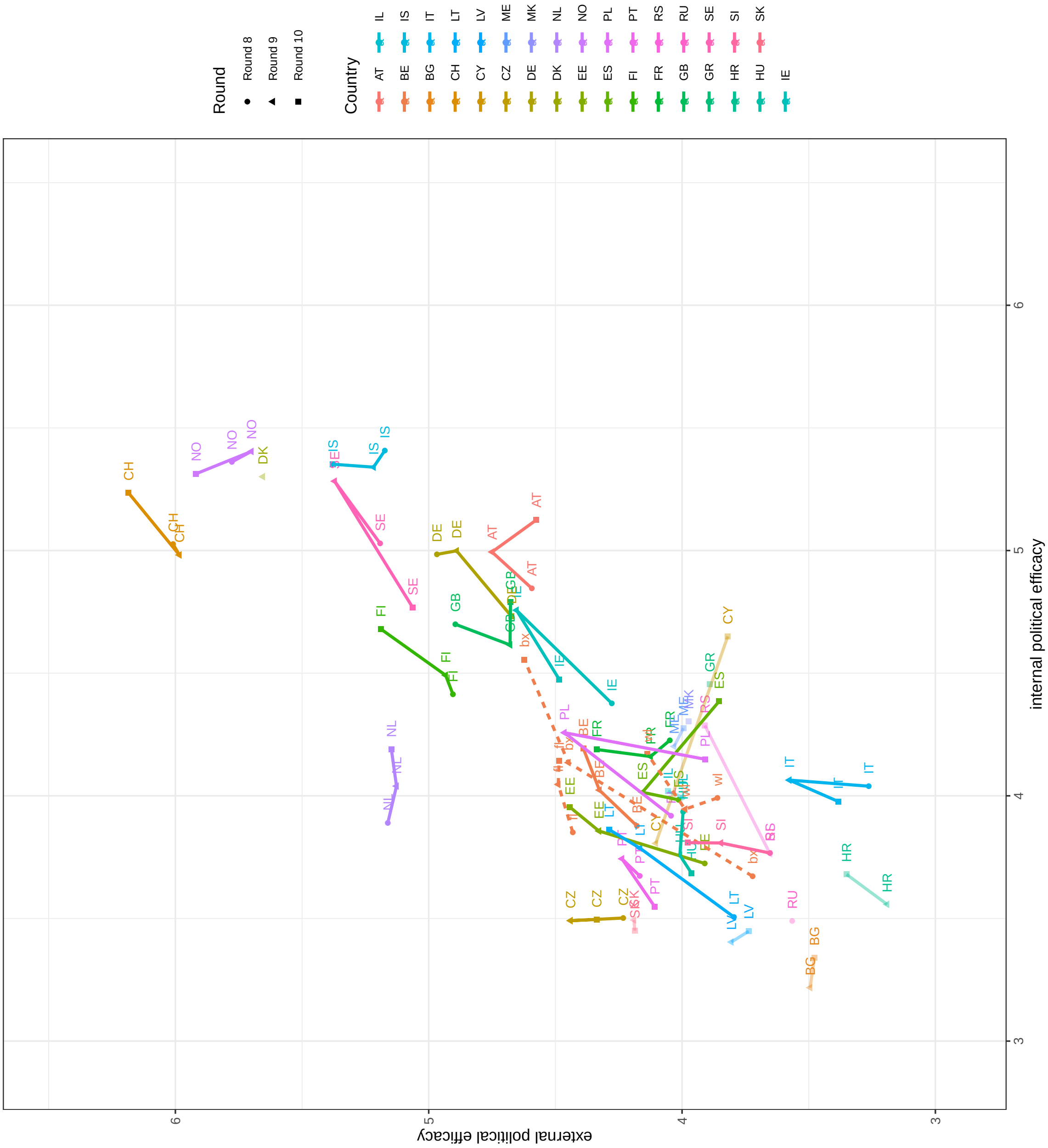


Figure 5.1: Means of internal and external political efficacy

The first observation that can be made is that the majority of the countries seems to be aligned on a main axis between internal and external efficacy with a minority of the countries having both an internal and external mean that is higher than 5. Furthermore, the changes between the three rounds seem to happen both on internal and external political efficacy. There is no common trend of those changes between all the countries. However, some groups seem to have similar patterns. For example, a few countries have an almost straight line between the three observation times such as Lithuania, Netherlands, Finland, etc. However, the sense of the evolution between rounds 8 and 10 is not necessarily similar. Another group of observations regroups countries for which their observations represent a triangle in their evolution such as Sweden, Italy, Ireland, Iceland, etc. In this group, it is the round 9 that seems to distinguish itself from the other two. Again, this distinction happens also in two directions, with either the round 9 having a higher or lower level of political efficacy than the two other rounds.

Looking at figure 5.2, a few countries seem to stand out for different reasons. The first country is the Netherlands (ellipse 2). This country seems to have a level of external political efficacy that is higher than a majority of countries. However, in terms of internal political efficacy, it does not stand out compared to the other countries. Looking at the evolution between the three rounds for the Netherlands, it appears that while the level of internal political efficacy has gotten much higher since the round 8, this trend does not appear on the external axis of the graph. Indeed, in terms of external political efficacy, the difference between the three rounds is quite small.

The next countries to stand out are Italy and Croatia (ellipse 3). They are the two countries with one of the smallest values of external efficacy. However, in terms of internal political efficacy, they do not stand out much from the other countries. Italy's level of external political efficacy has risen between the round 8 and 9 and decreased between round 9 and 10. In contrast, the level of internal political efficacy of Italy had a small decline between round 8 and 10.

Finally, Norway and Switzerland (ellipse 1) seem to be the two countries with the highest value of political efficacy on both dimensions. The two countries stand out as they are both in the upper right corner of the graph with Denmark.

Having described a few important trends of the graph, our attention is now dedicated to Belgium (figure 5.3). First of all, it appears that a difference exists between the Flemish and the Walloons as well as the Brussels residents. The main difference seems to be in terms of external political efficacy, the Flemish expressing a higher level than the Walloons. In terms of internal political efficacy, both groups of Belgians appear to have a similar level. However, the trend between rounds is different on both dimensions. The Flemish means of internal political efficacy has

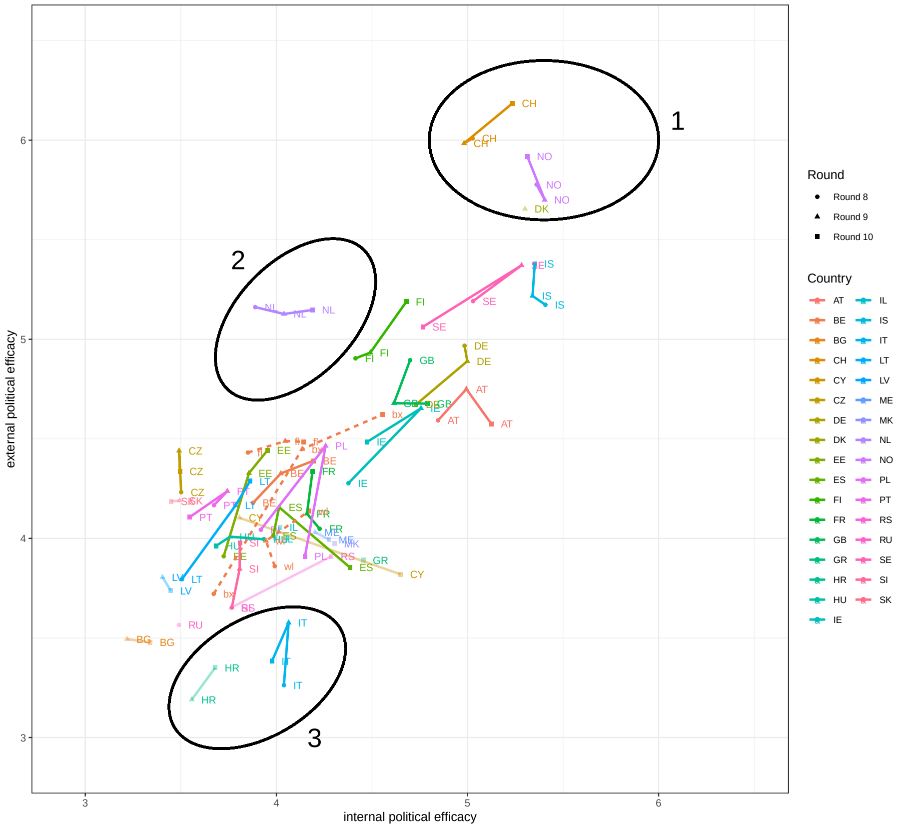


Figure 5.2: Means of internal and external political efficacy (bis)

risen between the round 8 and 10. On the Flanders bar-chart, it appears that for iconf and irol (internal dimension) the percentage of citizens answering 4 or 5 (corresponding to highest level of ability and confidence to participate in politics) to the two questions has risen between the round 8 and 10. For Wallonia, the mean has decreased between round 8 and 9 before rising between the round 9 and 10. On the Wallonia bar-chart, it appears that for the round 9, the percentage of people who answered 1 on the iconf variable is larger than for the two other rounds. External political efficacy has stayed constant for the Flemish whereas it has grown for the Walloons. Brussel seems to stand out as it starts in round 8 with a level of internal and external political efficacy that is lower than both the Flemish and the Walloon and finish in round 10 with a higher level on both dimensions than the other two groups. Finally, the overall Belgian evolution seems to be the average between the Walloons and the Flemish and does not represent the Brussel's trend. This might be explained by the sample size of Brussels which is quite smaller than the other two.

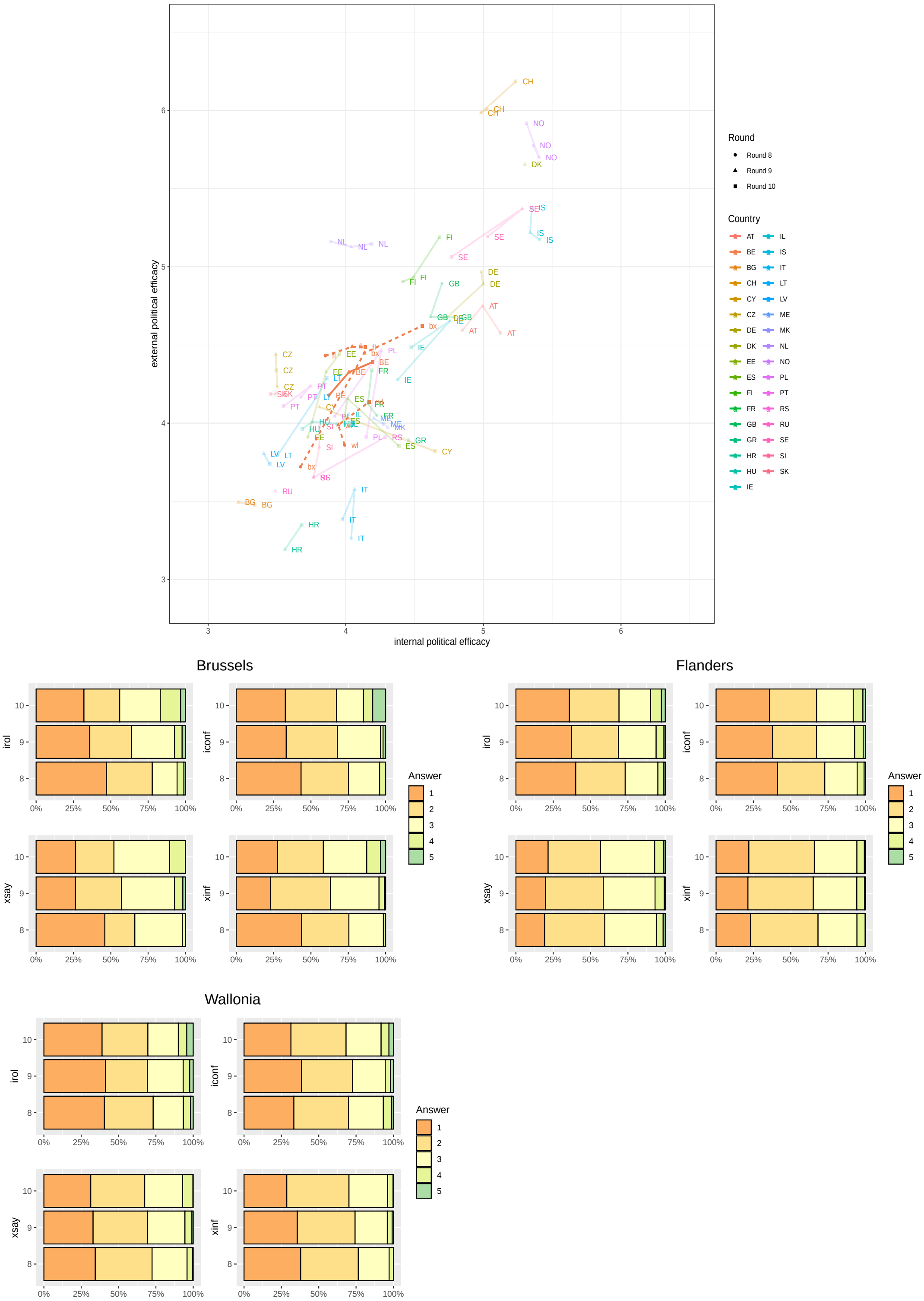


Figure 5.3: Belgium

5.2 Principal Component Analysis

This section presents the results of the PCA analysis. The PCA has been applied on the three waves of data and the countries not participating in all the rounds have been added as supplementary observations in the analysis. The PCA is used as a descriptive tool to see patterns between countries and rounds in terms of political efficacy by summarising the information contained in the data in a few principal components. The data matrix has been transformed to apply PCA on it. For each of the items of the ESS questionnaire, the item has been transformed in five new sub-items. These items represent the percentage of observations in a certain category of a certain variable. The observations are thus the countries. In total, there are 20 variables. To illustrate this new data matrix, an excerpt is shown in figure 5.4. To understand to which items of the questionnaire each variable comes from, the name of the variable needs to be decomposed. Firstly, the part before the underscore is the name of the variable and secondly the number after the underscore corresponds to the answer category. For example, the variable irol_2 corresponds to the second category of answers for the variable irol.

	irol_1	irol_2	irol_3	irol_4	irol_5	iconf_1	iconf_2	iconf_3	iconf_4	iconf_5	xsay_1	xsay_2	xsay_3
AT_10	17.32	40.78	28.92	9.76	3.22	8.53	35.72	31.94	15.99	7.82	18.14	42.21	33.01
AT_8	23.57	37.71	25.92	9.92	2.88	22.77	37.01	24.53	10.13	5.55	20.11	43.09	30.61
AT_9	26.21	34.47	25.49	9.06	4.78	22.81	32.87	23.94	11.61	8.76	17.82	39.29	34.80
BE_10	33.84	32.25	23.49	7.48	2.95	32.70	33.38	24.62	6.80	2.49	22.58	35.35	34.06
BE_8	38.76	33.30	22.03	4.61	1.31	37.22	33.64	22.99	5.24	0.91	25.44	37.68	31.70
BE_9	37.96	30.28	25.40	4.87	1.49	37.10	31.94	24.83	4.64	1.49	23.97	37.44	32.00
CH_10	22.49	32.53	28.13	11.00	5.85	11.49	29.50	34.66	17.68	6.67	7.02	16.37	34.80
CH_8	26.89	31.23	27.24	10.64	3.99	12.54	31.44	34.10	16.46	5.46	9.10	15.90	35.43
CH_9	28.11	34.24	23.10	10.09	4.45	12.94	31.80	32.08	17.68	5.50	8.63	16.28	36.46
CZ_10	48.98	28.41	17.52	4.07	1.02	55.58	26.64	13.89	3.01	0.88	26.15	30.84	32.92
CZ_8	46.83	32.90	16.15	3.15	0.97	54.33	29.62	12.86	2.59	0.60	26.28	35.31	29.85
CZ_9	45.42	34.70	16.24	2.02	1.61	55.27	30.00	11.46	2.39	0.87	20.57	35.62	34.38
DE_10	26.44	39.72	22.42	8.38	3.04	10.75	37.53	31.82	15.13	4.77	15.91	39.87	33.18
DE_8	22.05	36.41	27.40	10.26	3.88	9.44	34.16	33.95	17.31	5.13	14.21	35.48	37.87
DE_9	22.07	37.37	25.44	11.45	3.67	9.16	32.31	34.60	17.49	6.44	14.69	35.90	35.77
EE_10	37.26	35.69	18.93	6.02	2.10	41.39	33.99	18.40	4.39	1.83	22.53	38.18	31.96
EE_8	43.62	33.08	17.45	4.34	1.51	47.76	32.93	15.58	2.47	1.26	36.16	37.92	22.24
EE_9	40.48	35.23	16.60	5.89	1.80	45.20	33.79	15.76	3.66	1.59	24.77	38.51	29.66
ES_10	33.99	36.14	18.76	7.17	3.94	25.39	36.50	22.62	9.27	6.22	30.63	42.45	18.81

Figure 5.4: Data matrix

Figure 5.5 shows that the variables can be regrouped into 3 groups. The first one corresponds to variables that are negatively correlated with the first component and positively correlated with the second component (in orange on the graph). The four variables constituting this group are the second category of answer of each of the four political efficacy items of the ESS questionnaire. Inside this group, it seems possible to find two subgroups, one composed of the internal items (iconf_2 and irol_2) which is more positively correlated to the second dimension than the second group composed of external items (xinf_2 and xsay_2).

The second group of variables on the figure 5.5 corresponds to the variables that are negatively correlated with the first and second component (in dark red on the graph). This group is formed by the first category of answers for the variable xsay, xinf, irol and iconf. Once more the group of variables can be divided into two sub-groups. Here the first sub-group is constituted by irol_1 and iconf_1 which are almost perfectly correlated and have a higher negative correlation with the second component than the second group which is composed of xsay_1 and xinfl_1.

The last group of variables is made up of variables that are positively correlated with the first component. This last group includes the categories 3, 4 and 5 of the four items. This time the two subgroups that are observed can be discriminated based on the second component. Indeed, the internal efficacy items (iconf_3, iconf_4, iconf_5, irol_3, irol_4 and irol_5) are all positively correlated with the second component. The external efficacy items (xinf_4, xinf_5, xsay_3, xsay_4, xsay_5) are all negatively correlated with the second component except for the variable xinf_3.

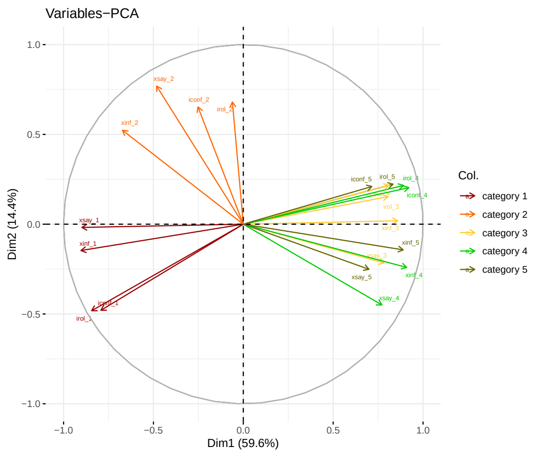


Figure 5.5: PCA variables graph

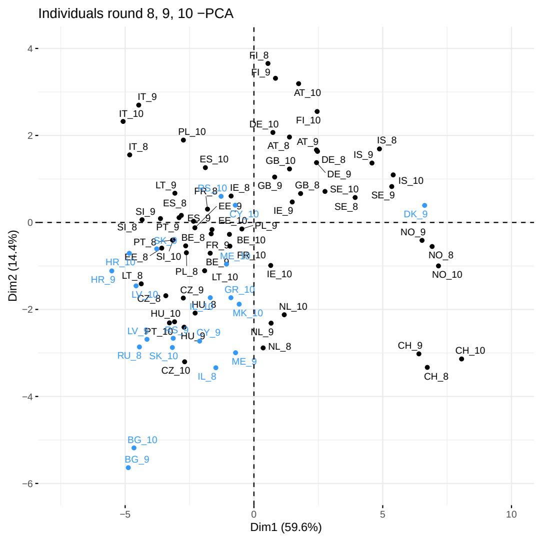


Figure 5.6: PCA individuals graph

Figure 5.6 represents the observations of all the countries for each round on the first and second component. To understand the label of a point, it is necessary to first look at the letter before the underscore that corresponds to the acronym of the country and then look at the number after the underscore to see to which round of the ESS belongs the observation. Furthermore, the observations in blue correspond to the countries that were not analysed in all the three rounds and thus are added as supplementary individuals.

A first observation that can be made looking at the graph is that a group of three countries appears clearly. This group can be characterised as having a positive coordinate on the first component (with the highest value on this component) and a negative coordinate on the second component. The individuals in this group are Switzerland (CH_8, CH_9, CH_10) and Norway (NO_8, NO_9, NO_10). Those countries can be characterised by a higher level of political efficacy, especially regarding the external political efficacy. In terms of longitudinal changes, it appears that for Switzerland, the difference between the different rounds seems to be more linked to the first component than the second. For Norway, difference between rounds appears to be linked to both dimensions.

A second observation is that Italy (IT_8, IT_9 and IT_10) seems to be very different from the other countries as it is characterized by positive coordinate on the second component and negative coordinate on the first component. This country seems to be characterised by the second category of answers for all variables. The position of Italy is maybe more determined by the second category of items related to external efficacy than internal efficacy. Thus, it would seem that Italians in general do not believe they have a high capacity of participating in politics and neither do they believe that the political system responds to the wishes of the citizens.

A third group of countries can be distinguished on the graph. This group is characterised by positive coordinates on both components. This group is composed of Germany (DE_8, DE_9 and DE_10), Austria (AT_8, AT_9 and AT_10), Finland (FI_8, FI_9, FI_10), Sweden (SE_8, SE_9 and SE_10), Great-Britain (GB_8, GB_9, GB_10) and Iceland (IS_8, IS_9 and IS_10). Here, in general, the differences between rounds of the different countries are both on the first and second component. This means that their position is determined by the highest category of internal political efficacy items.

Next, a group of individuals that have negative coordinates on both components can be observed. This group is composed of Hungary (HU_8, HU_9 and HU_10), Czechia (CZ_8, CZ_9 and CZ_10), Belgium (BE_8, BE_9 and BE_10). Furthermore, most of the supplementary individuals can be regrouped with those countries. Those countries' position is characterised by the lowest level of political efficacy.

Finally, for some countries the difference between rounds is quite large as some of their coordinates go from negative to positive for one component or two components. Those countries are Poland (PL_8, PL_9 and PL_10), Ireland (IE_8, IE_9 and IE_10) and Portugal (PT_8, PT_9 and PT_10). Poland's position in the round 8 and 9 is characterised by the first categories of answer in the different items whereas in the round 10, the position is more determined by the second category on all the items. Ireland's round 10 position is determined by the three highest categories of external political efficacy and the opposite happens for the 9^{th} round where the position of round 9 is explained by the highest categories of internal efficacy. Ireland 8^{th} round position is determined by the second category of answer of political efficacy items. Portugal position in the round 10 seems to be more influenced by the internal political efficacy whereas rounds 8 and 9 seem to be characterised by external political efficacy.

5.3 Clustering

This section is dedicated to a clustering analysis. The objective is to see how, based on their similarity in terms of political efficacy, the different countries in the different rounds can be regrouped. The idea is to see if a group of observations in terms of political efficacy can be discovered. To compute the cluster analysis, the data matrix has been transformed in the same way as for the PCA analysis (see figure 5.4).

To understand the label of an observation, it is necessary to first look at the letter before the underscore that corresponds to the acronym of the country and then look at the number after the underscore indicating to which round of the ESS corresponds the observation.

The results of the clustering analysis are presented in a dendrogram below (figure 5.7). It appears that the data can be divided into two clusters. The first cluster is composed of the three rounds of eleven countries and an additional country for two rounds. The cluster can be separated into two sub-clusters. The first sub-cluster contains Italy (round 8, 9, 10), Spain (round 10) and Poland (round 10). The second sub-cluster is composed of France (round 8, 9, 10), Ireland (round 9, 10), Poland (round 8, 9), Belgium (round 8, 9, 10), Estonia (round 8, 9, 10), Lithuania (round 8, 9, 10), Portugal (round 8, 9, 10), Spain (round 8, 9) Czechia (round 8, 9, 10), Slovenia (round 8, 9, 10) and Hungary (round 8, 9, 10). The second main cluster is composed of the three rounds of nine countries and an additional country for one round. This main cluster can also be divided into two sub-clusters. The first one is composed of Switzerland (round 8, 9, 10) and Norway (round 8, 9, 10), Sweden (round 8 and 9) and Iceland (round 8, 9 and 10). The second sub-cluster contains Netherlands (round 8, 9, 10), Finland (round 8, 9, 10), Germany (round 8, 9, 10), Sweden (round 10), Austria (round 8, 9, 10), Great-Britain (round 8, 9, 10) as well as Ireland (round 9).

This analysis does not really give us any indication about the measurement equivalence of political efficacy. However, the results bring out some noteworthy information that is interesting if the measurement invariance of the instrument is accepted. Indeed, the majority of the countries are in the same cluster for their three rounds. This means that they have similar patterns in all three rounds. In contrast, a few countries switches cluster depending on the rounds. This could show that a bigger shift in political efficacy happened between the rounds in those countries.

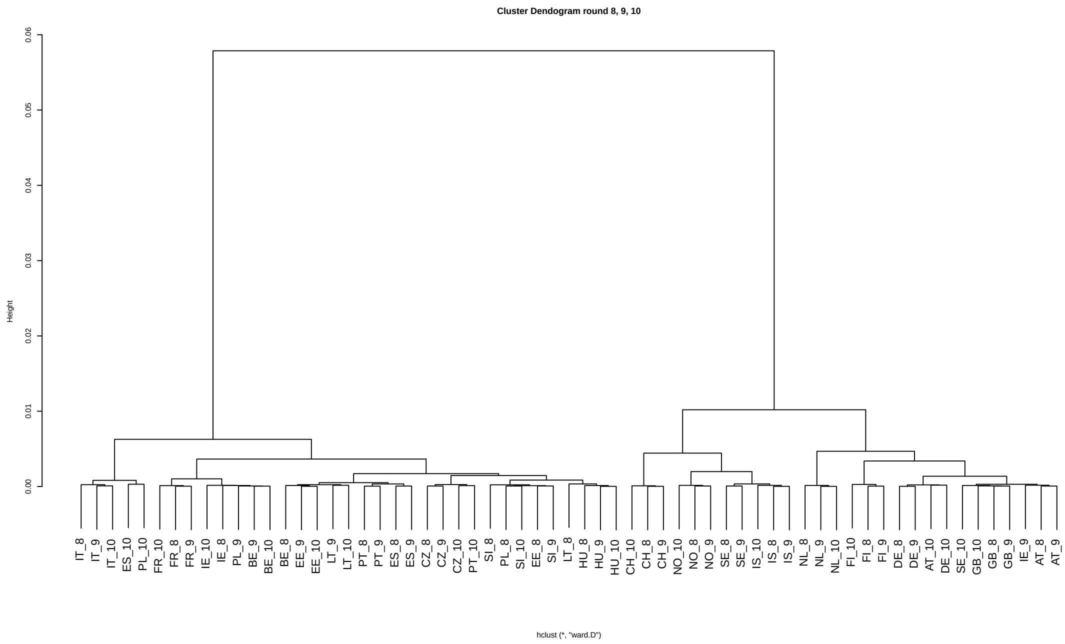


Figure 5.7: Clustering

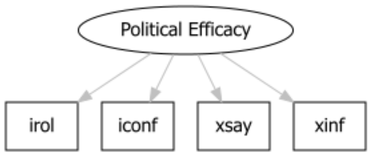
5.4 Multi-group CFA

This section is devoted to the results obtained by the multi-group CFA analysis. The different steps to measurement invariance will be followed. The measurement invariance will be tested for two elements, the first one is longitudinal invariance (invariance between rounds) and the second one is cross-country invariance (invariance between countries in one round). Each analysis will be produced using continuous and categorical estimation methods. Before checking the measurement invariance qualities of the instrument, a model will be selected.

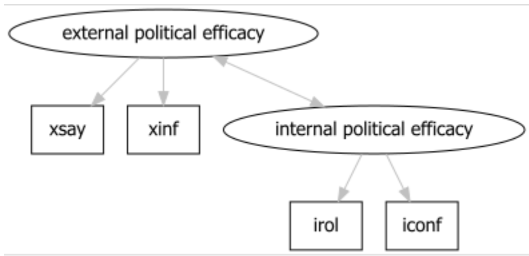
5.4.1 Model selection

Based on the theory of political efficacy and the ESS variables pertaining to it, the model proposed is a two factors model with each factor representing a dimension of political efficacy (internal and external). Each factor is composed of two observed indicators (iconf and irol for internal and xsay and xinf for external). However, in the literature, it is recommended to first test a one factor model. If the fit of the model is not good, the model should be made more complex by adding factors in the model. Thus, a first model with one factor was fitted and then a second model with two factor was fitted to the whole database. The quality of the fit of both models has been evaluated based on the RMSEA and CIF index. The model who performed the best was selected to conduct the measurement invariance tests.

The first model (figure 5.8a) represents a measurement model where the indicators irol, iconf, xsay and xinf are all related to one factor, political efficacy. This model does not distinguish between internal and external dimensions of political efficacy. The second model (figure 5.8b) is a two factors model. In this case, the indicators irol and iconf are related to the factor internal political efficacy and the indicators xsay and xinf are linked to the external political efficacy factor.



(a) One factor model



(b) Two factors model

Figure 5.8: CFA models

Table 5.1 shows the two models fit index. As expected, the analysis confirms clearly that a one factor model is not suitable and that a two factor model agrees much better with the data. For the rest of the analysis, only the second model will be considered.

	Continuous		Categorical	
	RMSEA	CFI	RMSEA	CFI
Model 1	0.222	0.676	0.280	0.896
Model 2	0.003	1.000	0.007	1.000

Table 5.1: Fit of the models

5.4.2 Model convergence

As explained in the methodology of CFA, it is recommended to have at least 3 indicators by factor to avoid empirical under-identification (Roos & Bauldry, 2022). However, the ESS recognised in the questionnaire only four items to measure political efficacy. A principal component analysis has been applied on the questionnaire to search for other items that would measure this concept. However, no variable has been identified. Thus, as the model that works best has two factors with each two indicators, there is a risk of under-identification in our data. The model estimated with a method which considers the data as continuous is less complex than the model estimated with a method that takes into account the ordinal property of the data. This is the reason why the two factor model will be first estimated with maximum likelihood (ML) to check if the model can be identified and then if no problem is discovered, the model will be estimated with the robust diagonally weighted least square (WLS) estimator. The second reason for the use of the ML estimator is that the representation of the invariance or non-invariance of the data is easier with a continuous estimation because there is a linear relationship between the indicators and the factors which is not the case in the categorical estimation.

Before conducting the measurement invariance analysis, the model has been run for every sample (corresponding to each country in each round) both with continuous and categorical estimations. For some countries, some estimation issues were encountered such as either the model did not converge or the model estimated negative variance or covariance parameters. It appears that with the continuous estimation, three countries had estimation issues: Germany (round 8 & 10), Spain (round 10) and Iceland (round 10). With the categorical estimation, a higher number of countries obtained estimation issues: Austria (round 10), Germany (round 8, 9, 10), Estonia (round 9), Spain (round 9 and 10), Ireland (round 9), Iceland (round 9 & 10) and Poland (round 10).

5.4.3 Longitudinal measurement invariance

The first measurement invariance of the data that is investigated is longitudinal invariance. The objective is to check that the results of a country can be compared between rounds. The longitudinal invariance has been checked for the countries that did not have estimation issues and that were present in all three rounds. The different steps have been applied to conduct the measurement invariance analysis. The freely estimated model is fitted for every round by countries to verify the configural invariance of the rounds. If the configural invariance is accepted, the multi-group CFA model is computed for every country. This first model will be called the configural model because no equality constraints will be imposed on the parameters. Then the metric model will be fitted. This model will constraint the loadings to be equal across rounds. Once the configural and metric model have been fitted, the metric invariance is evaluated by looking at the difference between the fit index of those two models. The metric invariance hypothesis is rejected if both indices reject the invariance. Lastly, the scalar model will be fitted. This model will put equality constraints on the loadings and intercepts/thresholds. Scalar invariance will then be analysed by checking the difference between the fit indices of the metric and scalar models. As for metric invariance, scalar invariance is rejected if the two indices find non-invariance.

Continuous estimation

Before looking at the fit index of the continuous estimated CFA model, the distribution of the variables has been looked at. An histogram has been designed for each country in each round (see appendix B figures B1 - B3). It appears that in general the data does not follow a normal distribution. The lack of normality will be corrected in the estimation by adding the Satorra-Bentler correction to the maximum likelihood estimation.

The CFA has been applied to every country and every round except Germany, Spain and Iceland. A summary of the quality of the fit is presented in tables 5.2 and 5.3 (see appendix B tables B1 & B2 for values of each model). Starting with the RMSEA index, It appears that for a majority of the countries, the model obtains a good fit (meaning that the RMSEA < 0.05). A few countries had an acceptable fit (RMSEA between 0.05 & 0.08) such as Czechia (round 8), France (round 8), Switzerland (round 9), Finland (round 9), Lithuania (round 9) and Slovenia (round 9). No countries achieved a bad fit (RMSEA > 0.08). Looking at the CFI table, it appears that for all rounds, every country achieved a good fit (meaning that CFA > 0.99) except Lithuania (round 9) which obtained an acceptable fit with a CFI between 0.95 and 0.99). As in the RMSEA, none got bad fit (CFI < 0.95). Those first results show no serious lack of configural invariance.

Table 5.2: Configural RMSEA

	Round 8	Round 9	Round 10
Good fit	AT, BE, CH, EE, FI, GB, HU, IE, IT, LT, NL, NO, PL, PT, SE, SI	AT, BE, CZ, EE, FR, GB, HU, IE, IT, NL, NO, PL, PT, SE	AT, BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, LT, NL, NO, PL, PT, SE,SI
Acceptable fit	CZ, FR	CH, FI, LT, SI	
Bad fit			

Table 5.3: Configural CFI

	Round 8	Round 9	Round 10
Good fit	AT, BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, LT, NL, NO, PL, PT, SE, SI	AT, BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, NL, NO, PL, PT, SE, SI	AT, BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, LT, NL, NO, PL, PT, SE,SI
Acceptable fit		LT	
Bad fit			

Metric invariance is then analysed for every country. The configural multi-group CFA is first estimated, followed by the metric model. A serious lack of metric invariance was not detected in any country (see appendix B tables B3 & B4). For the RMSEA (table 5.4), Estonia has a ΔRMSEA that is larger than 0.01. Looking at the CFI (table 5.4), all the countries obtained a ΔCFI that is bigger than -0.01. Thus, it can be concluded that there is no evidence of lack of metric invariance in the 18 countries in terms of longitudinal measurement.

Table 5.4: Metric invariance: RMSEA

	Countries
Invariance	AT, BE, CH, CZ, FI, FR, GB, HU, IE, IT, LT, NL, NO, PL, PT, SE, SI
Non-invariance	EE

Table 5.5: Metric invariance: CFI

	Countries
Invariance	AT, BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, LT,NL, NO, PL, PT, SE, SI
Non-invariance	

Having achieved metric invariance, the scalar invariance can now be investigated. Tables 5.6 and 5.7 summarize the results. Starting with the RMESA, there are 7 countries for which the RMSEA rejects the invariance hypothesis (ΔRMSEA $> 0,01$): Austria, Estonia, Great-Britain, Ireland, Norway, Poland and Sweden. In terms of CFI, the countries that are problematic (country with a Δ CFI < -0.01) are Austria and Poland. It appears that Austria and Poland obtained both a ΔRMSEA that is larger than $0,01$ and ΔCFI that are smaller than -0.01 which means that scalar invariance is rejected. For the rest of the countries, the fit indices do not indicate a problem of scalar invariance (appendix B B3 & B4). As there is only three rounds, it is easily possible to do two by two analysis in order to identify between which rounds there is a problem of invariance. This is done for Poland and Austria.

Table 5.6: Scalar invariance: RMSEA

	Countries
Invariance	BE, CH, CZ, FI, FR, HU,IT, LT, NL, PT, SI
Non-invariance	**AT,** EE, GB, IE, NO, **PL**, SE

Table 5.7: Scalar invariance: CFI

	Countries
Invariance	BE, CH, CZ, EE, FI, FR, GB, HU, IE, IT, LT, NL,NO, PT, SE, SI
Non-invariance	**AT, PL**

Three new longitudinal CFA have been applied to Poland and Austria:

- With rounds 8 and 9
- With rounds 9 and 10
- With rounds 8 and 10

The results for Poland (see table B5 in the appendix B) show that the CFA with rounds 8 and 9 achieved scalar invariance whereas, the CFA with rounds 9 and 10 and the one with rounds 8 and 10 reached metric invariance and scalar invariance was rejected. The results of Austria are different (see table B6 in the appendix B). In this case, a serious lack of scalar invariance appears only in the analysis considering the rounds 8 and 10.

Finally, it is possible to estimate the latent means with the scalar models. The reference group in the model will be the round 8. This means that the multi-group CFA analysis will estimate the latent means of the round 8 to be equal to 0. Then, the CFA will estimate the difference between this round and the two others to compute the estimation of the latent means of rounds 9 and 10. Once these are computed, the CFA analysis in R produces a z-score test to test whether the estimation of the group is statistically different from 0 (except for the reference group, here round 8). This can be understood as to test whether the means of rounds 9 and 10 are statistically different from 0. In the appendix B, the table B7 gives the p-value of all the tests. In the table 5.8 below, this information has been summarised. It presents the possibilities of having, for each dimension, a significant difference ($\alpha < 0.05$) between the means of round 8 and round 9 and of round 8 and round 10.

The results indicate that the differences between the rounds on both dimensions are significant for Belgium, Estonia, Lithuania and Sweden. Great-Britain, Italy and Slovenia have all significant differences for the external dimension but not on the internal dimension. It is the contrary for Hungary and Netherlands. Norway and Portugal do have non-significant differences on both dimensions for every round. Czechia's only significant difference is between rounds 8 and 9 on external political efficacy. Finland has a significant difference on both dimensions between the rounds 8 and round 9. Ireland only non-significant difference is between round 8 and 10 on the internal dimension and Switzerland only non-significant difference is between rounds 8 and 9 on the internal dimension. France has a significant difference between rounds 8 and 10 on the external dimension and a significant difference between rounds 8 and 9 on the internal dimension.

Table 5.8: Mean differences between rounds

Extern\Intern	$\mu_8 = \mu_9$ $\mu_8 = \mu_{10}$	$\mu_8 \neq \mu_9$ $\mu_8 = \mu_{10}$	$\mu_8 = \mu_9$ $\mu_8 \neq \mu_{10}$	$\mu_8 \neq \mu_9$ $\mu_8 \neq \mu_{10}$
$\mu_8 = \mu_9$ $\mu_8 = \mu_{10}$	NO, PT			HU, NL
$\mu_8 \neq \mu_9$ $\mu_8 = \mu_{10}$	CZ			
$\mu_8 = \mu_9$ $\mu_8 \neq \mu_{10}$		FR	FI	
$\mu_8 \neq \mu_9$ $\mu_8 \neq \mu_{10}$	GB, IT, SI	IE	CH	BE, EE, LT, SE

Categorical estimation

The same steps as for the continuous estimation have been applied. Starting with the configural equivalence, the model has been separately fitted to every country and every round (see tables B7 & B8 in appendix B). In terms of RMSEA (table 5.9), the majority of the countries obtained a good fit (RMSEA < 0.05) in all three rounds. A few countries achieved an acceptable fit (RMSEA between 0.05 and 0.08): Switzerland (round 8 & 9), Finland (round 8 & 10), France (round 8), Italy (round 10), Netherlands (round 8) and Slovenia (round 8 & 10). Only three country obtained a bad fit (RMSEA> 0.08): Czechia (round 8), Finland (round 9) and Lithuania (round 9). In contrast, for the CFI index (table 5.10), every country achieved a good fit in every round. Those results show no serious lack of configural invariance.

Table 5.9: Configural RMSEA

	Round 8	Round 9	Round 10
Good fit	BE, GB, HU, IT, LT, NO, PT, SE, SI	BE, CZ, FR, GB, HU, IT, NL, NO, PT, SE	BE, CH, CZ, FR, GB, HU, LT, NL, NO, PT, SE
Acceptable fit	CH, FI, FR, NL	CH, SI	FI, IT, SI
Bad fit	CZ	FI, LT	

Table 5.10: Configural CFI

	Round 8	Round 9	Round 10
Good fit	BE, CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, PT, SE, SI	BE, CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, PT, SE, SI	BE, CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, PT, SE, SI
Acceptable fit			
Bad fit			

Metric invariance can now be investigated (see tables B10 & B11 in appendix B). Starting with the RMSEA index (table 5.11), Hungary and Sweden have a Δ RMSEA that is problematic ($\Delta RMSEA > 0.1$). In contrast, for the CFI all the countries respond to the invariance demands (table 5.12). Thus, longitudinal metric invariance is achieved for the 11 countries.

Table 5.11: Metric invariance: RMSEA

	Countries
Invariance	BE, CH, CZ, FI, FR, GB, IT, LT, NL, NO, PT, SI
Non-invariance	HU, SE

Table 5.12: Metric invariance: CFI

	Countries
Invariance	BE, CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, PT, SE, SI
Non-invariance	

As metric invariance is accepted, scalar invariance can be analysed (see tables B10 & B11 in appendix B). It appears that Belgium and Sweden have a $\Delta RMSEA > 0.01$ (table 5.13). All countries obtain a ΔCFI that is acceptable (table 5.14). Those results means that a serious lack of scalar invariance has not been detected for any country.

Table 5.13: Scalar invariance: RMSEA

	Countries
Invariance	CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, SI
Non-invariance	BE, PT, SE

Table 5.14: Scalar invariance: CFI

	Countries
Invariance	BE, CH, CZ, FI, FR, GB, HU, IT, LT, NL, NO, PT, SE, SI
Non-invariance	

Finally, the latent means have been estimated based on the scalar model. The table 5.15 below summarises the results of mean differences (see appendix B12 for p-values). Belgium, Lithuania and Sweden obtained significantly different means on both dimensions and for all the rounds. It is the contrary for Norway which has no significant difference on either dimension. Czechia and France and Portugal have only one significant difference: Czechia between rounds 8 and 9 on the external dimension, France between rounds 8 and 10 on the external dimension and Portugal between rounds 8 and 10 on the internal dimension. Switzerland and Finland both have significant differences between the rounds 8 and 10 on both dimensions of political efficacy. Finally, Hungary and Netherlands only have significant difference between rounds on the internal dimesion and it is the opposite for Great-Britain, Italy and Slovenia.

Table 5.15: Mean differences between rounds

Extern\Intern	$\mu_8 = \mu_9$ $\mu_8 = \mu_{10}$	$\mu_8 \neq \mu_9$ $\mu_8 = \mu_{10}$	$\mu_8 = \mu_9$ $\mu_8 \neq \mu_{10}$	$\mu_8 \neq \mu_9$ $\mu_8 \neq \mu_{10}$
$\mu_8 = \mu_9$ $\mu_8 = \mu_{10}$	NO		PT	HU, NL
$\mu_8 \neq \mu_9$ $\mu_8 = \mu_{10}$	CZ			
$\mu_8 = \mu_9$ $\mu_8 \neq \mu_{10}$	FR		CH, FI	
$\mu_8 \neq \mu_9$ $\mu_8 \neq \mu_{10}$	GB, IT, SI			BE, LT, SE

5.4.4 Cross-national measurement invariance

In this section, the measurement invariance between European countries will be tested. The same strategy as the one developed in the longitudinal section will be applied. To test configural invariance in the cross-national analysis, a configural model needs to be fitted to every country of one round and repeated for every round. This has already been done previously. Configural invariance is accepted based on the results of the longitudinal analysis in the previous section (see tables

5.2 and 5.3). Furthermore, to conduct the country invariance, three multi-group CFA analyses will be produced one for each round. Again, the analysis will be conducted for continuous and categorical estimations.

Continuous estimation

First of all, the multi-group CFA contains 18 countries (the ones without any estimation issues). As explained before, the configural invariance is accepted for all three rounds. Next, looking at the metric invariance, it appears that in terms of RMSEA (table 5.16), all the rounds obtained a ΔRMSEA smaller or equal to 0.01. In terms of CFI (table 5.17), no countries achieved a Δ CFI smaller than -0.01. As both indices are under the thresholds, there is no serious lack of metric invariance for any round.

As the metric invariance assumptions are accepted, the scalar invariance can then be analysed. In terms of RMSEA (table 5.16), all three rounds possess a value that is larger than 0.01. Moreover, for the CFI (table 5.17), all three rounds also possess a value that is smaller than -0.01. Thus, it appears that there is a serious lack of scalar invariance between the countries for all three rounds.

Table 5.16: RMSEA

	Configural	Metric	Scalar	Δ_{metric}	Δ_{scalar}
Round 8	0.031	0.036	0.082	0.005	0.046
Round 9	0.021	0.031	0.081	0.010	0.050
Round 10	0.031	0.036	0.086	0.005	0.050

Table 5.17: CFI

	Configural	Metric	Scalar	Δ_{metric}	Δ_{scalar}
Round 8	0.998	0.994	0.948	-0.004	-0.046
Round 9	0.999	0.996	0.951	-0.003	-0.045
Round 10	0.998	0.994	0.946	-0.004	-0.048

The conclusion indicates that as scalar invariance is not reached, it is not possible to compare directly the latent means of the different countries. However, this result does not indicate the size of the invariance. To explore the level of scalar invariance between the countries, a representation of the invariance has been proposed. Firstly, the latent scores of each observation have been computed (based on the metric model). Then, a scatter plot (figure 5.9) for each of the four indicators of the model has been drawn: y axis representing the indicator and the x axis representing the factor. The graphs only show a regression line for each country and not the observations.

The first observation that can be made is that a similar pattern happens in the three rounds. Furthermore, in all the rounds, the majority of the countries obtains regression lines that are quite close. It would seem that the distances between the regression lines are not very important. It seems that the difference between countries is not very large. Thus, the CFA analysis has concluded that the scalar invariance is not reached but that the size of the invariance does not seem very important.

Categorical estimation

As for the continuous estimation, the configural invariance is accepted based on the results of the longitudinal analysis. The next step is to study the metric invariance. Tables 5.18 and 5.19 show that for the RMSEA and the CFI, none of the rounds displays a lack of metric invariance. This leads to the conclusion that for each round, the instrument is metric invariant.

Now, scalar invariance can be analysed. It appears that for the RMSEA (table 5.18), all the rounds obtain a Δ RMSEA that is larger than 0.01 and for the CFI (table 5.19), the three rounds also got a ΔCFI that is smaller than -0.01. Those results show a lack of scalar invariance of the instrument.

Table 5.18: RMSEA

	Configural	Metric	Scalar	Δ_{metric}	Δ_{scalar}
Round 8	0.043	0.045	0.092	0.002	0.047
Round 9	0.046	0.043	0.090	-0.003	0.047
Round 10	0.043	0.045	0.083	0.002	0.038

Table 5.19: CFI

	Configural	Metric	Scalar	Δ_{metric}	Δ_{scalar}
Round 8	1.000	0.998	0.973	-0.002	-0.025
Round 9	0.999	0.999	0.972	0.000	-0.027
Round 10	1.000	0.998	0.977	-0.002	-0.021

In this case, representing the lack of invariance as it has been done in the continuous case is not done because in the categorical estimation, there is no direct linear relationship between the indicators and the factors and thus there is no certainty that the graph would represent the lack of invariance.

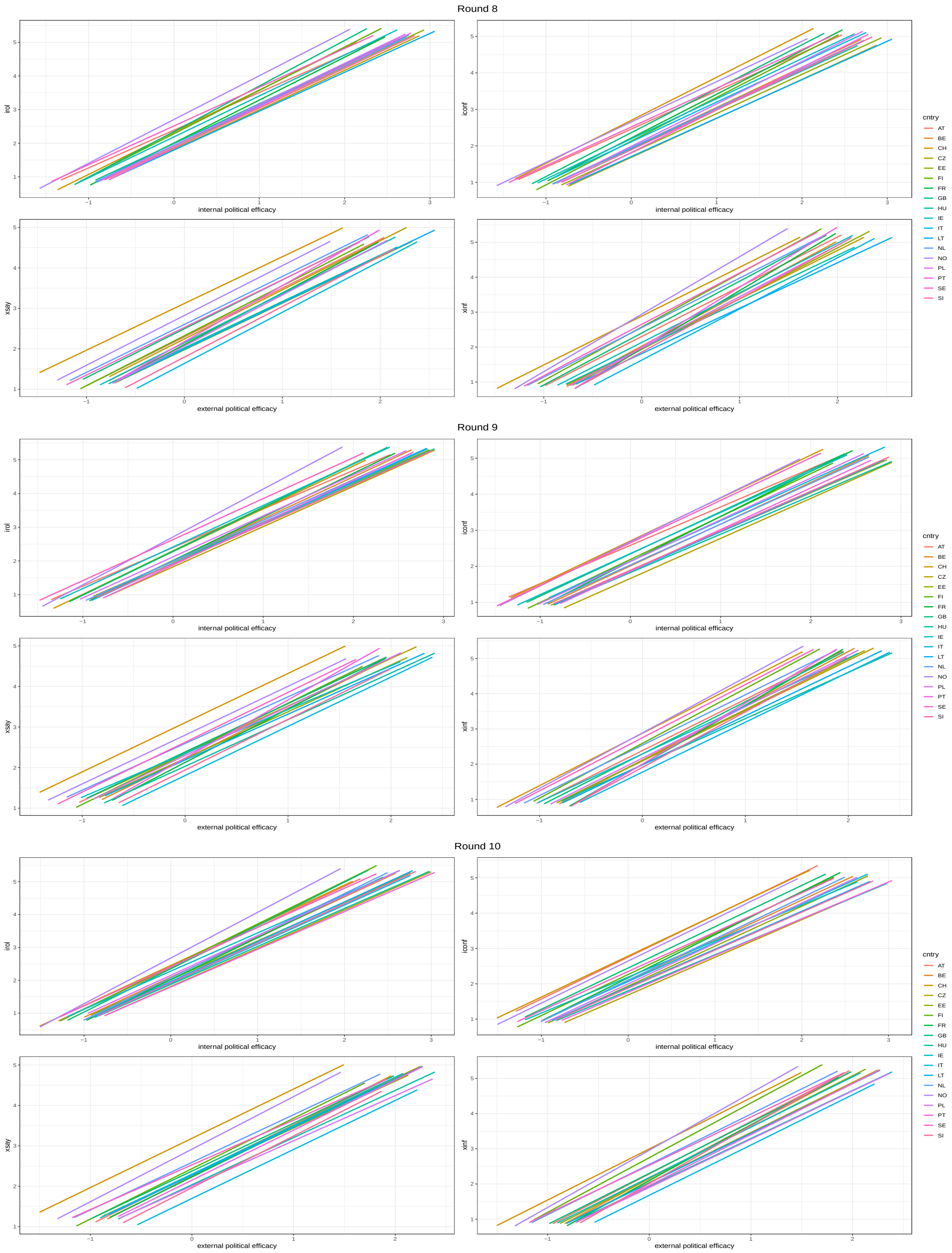


Figure 5.9: Invariance representation

Chapter 6

Analysis of the Results

This chapter is devoted to the discussion of the major findings of the previous chapter. The first section depicts the main conclusions regarding the measurement invariance of the political efficacy concept in the ESS, focusing on both longitudinal and cross-national analyses. The following section provides a discussion about the scalar invariance in the longitudinal analysis. The next section tackles potential reasons for the observed lack of invariance in some analyses. Then, the chapter illustrates the lack of invariance between countries within a single round of the survey. Finally, the chapter concludes by discussing potential variables that could explain the bias in the political efficacy scale.

6.1 Measurement invariance

The multi-group CFA analysis has revealed major results for the measurement equivalence of the political efficacy concept in the ESS. The first result pertains to the configural and metric level of invariance tested. In the country longitudinal analyses (comparing different rounds within each country) and cross-national analyses (comparing different countries within each round), there was no indication of a lack of configural and metric invariance in the data. This indicates that political efficacy is understood consistently in every country and round and that the relationships between the indicators and factors are the same across the different countries and rounds. This result shows that some comparisons between the different cultural groups can be produced.

Having discovered metric invariance guarantees two results for the ESS. Firstly, configural invariance of the data indicates that the four items of the questionnaire do measure the same two latent variables which are internal and external political efficacy in the different countries (Kankaraš & Moors, 2010). Secondly, metric invariance implies that the same relationships exist between the different observed variables and latent variables in all of the different countries. Those two conclusions mean that in the ESS, certain types of comparisons can be made between the countries. Indeed, this means that you can compare for example the differences between men and women in terms of political efficacy in the different countries.

Another major result concerns the scalar invariance analysis. This time, the results are different between the longitudinal analyses and the cross-national analyses. Starting with the longitudinal analyses, there was no evidence of a lack of scalar invariance for a majority of countries which means that it is valid to study the evolution of political efficacy through the rounds for a country. The only two countries that rejected scalar invariance were Poland and Austria. Regarding the cross-national comparison, scalar invariance was not reached in any round. This means that direct comparison of latent scores need to be carefully analysed, taking potential biases due to lack of scalar invariance into account. This is attempted in section 6.4.

To understand what it means to have a metric invariant scale but not scalar invariant, an illustration is given through a well known example: the temperature (Herk & Goldman, 2022). It can be measured through different scales: the Celsius, Fahrenheit and Kelvin scale. The three scales measure the same phenomenon known as the temperature. However, the three scales express the measure of temperature differently. The Celsius and Fahrenheit scales are configural invariant. They both measure the temperature but they have a different scale both in terms of loadings and intercepts as to obtain Fahrenheit in Celsius you have to divide by a number and then add to it 32. In contrast, Celsius and Kelvin scale are metric invariant because to convert Celsius to Kelvin you have to add 273. This illustration shows that with metric invariance scale, it is still possible to compare the groups but the comparison cannot be made based on the latent means.

6.2 Longitudinal scalar invariance

The analysis where the measurement invariance was investigated between rounds for each country (longitudinal analysis) has demonstrated that for most countries it is possible to directly compare the latent score. This analysis has demonstrated that the ESS objective to collect longitudinal data to study the evolution of a concept over time is achieved in the majority of the countries tested for political efficacy. Furthermore, this means that on the descriptive graph a comparison between rounds for each country (except Poland and Austria) can be analysed. Furthermore, once scalar invariance could not be rejected, the latent means could be computed with the CFA. Once those latent means have been computed, the analysis of the significance of the differences between the means of rounds 9 and 10 and the mean of the round 8 could be conducted. For example, in the descriptive graph section, attention has been given to Netherlands for their changes on one specific dimension and a constant level on the other dimension. With the CFA analysis, it has been shown that the Netherlands difference in internal political efficacy is significant but not on the external dimension. This enlightens us on the fact that between round 8 and round 10, citizen's evaluation of his competency to participate has increased whereas the evaluation of the system has been stable. Furthermore, for Norway, none of the differences is significant. The small differences in efficacity in Norway may simply due to sampling variability.

6.3 Bias in the instrument

Having discovered a lack of scalar invariance in Poland and Austria between the rounds 8, 9 and 10 as well as a lack of scalar invariance between the countries inside a round, possible reasons to explain this lack of invariance are exposed.

Concerning the longitudinal analysis of each country, in the continuous estimation, Austria and Poland were not scalar invariant. Further analyses on Poland have shown that for the rounds 8 and 9 analysis, there was no evidence of a lack of scalar invariance. It was the 10^{th} wave of the ESS that was non-scalar invariant with the two other rounds. To understand this lack of invariance, a few possible biases have been considered. The construct bias cannot be considered as configural and metric invariance were reached which implied that at least the same construct is measured with the different indicators whereas construct bias would imply that the construct studied is not understood in the same way in the different rounds (Welkenhuysen-Gybels et al., 2007). However, method bias (Byrne & Watkins, 2003) should be considered as it is known that the round 10 had a different data collection method than the previous round. In Poland, the data in the round 10 was collected by self-completion questionnaires (either online or paper) whereas for the rounds 8 and 9, surveys were administered by face-to-face interviews. Finally, the last possible bias is item bias (van de Vijver & Leung, 1997). In this case, a future analysis should be made to see which items out of the four would explain the non-invariance of the instrument.

The situation for Austria is different. Between the round 8-9, both the Δ_{scalar} of the CFI and RMSEA were below their thresholds indicating a good level of invariance. Between the rounds 9-10, both the CFI and RMSEA were near or above the thresholds, suggesting a lesser degree of invariance. Lastly, between the rounds 8-10, both indices were above the thresholds indicating a lack of invariance.

This results makes it harder to propose a possible source of bias because it is not one round in particular that differs from the other two.

Looking at the possible bias of the political efficacy scale, the construct bias hypothesis should not be considered. This hypothesis is not probable because there is construct bias when the scale does not measure the same concept in the different samples. As configural and metric invariance are reached, construct bias hypothesis should be discarded. Unlike construct bias, method bias should be considered for two mains reasons. The first one is that the survey is translated into different languages. As the different countries do not have the same national language, the difference in results could be linked to translation (Byrne & Watkins, 2003). The second reason is mainly concerned with the results of the round 10. This reason concerns data collection. With Covid, the data collection process has not been the same in all the countries with some countries doing the majority of the interviews face-to-face whereas in other countries the majority of the data was collected through self-completion questionnaire. Lastly, item bias could also be a reasonable explanation especially as the number of groups to compare is quite high, it is possible that some items would not lead people with similar latent characteristics to the same answer in one of the observed variables (van de Vijver & Leung, 1997).

6.4 Cross-national lack of invariance

As scalar invariance was not reached in the cross-national analyses, you cannot directly compare the mean of each country between each other in one round. Scalar invariance guarantees that the instrument has the same origin. If it is not reached, it is possible that the differences observed between countries are not true differences but artificial differences that would come from a bias in the measurement instrument.

This finding means that the descriptive results should be carefully examined to avoid interpreting differences as true differences when they could be the results of a bias in the questionnaire. In this paper, to quantify the lack of invariance, it has been proposed to represent graphically the relationship between the latent score of each factor and each indicator. The idea is that by looking at both the descriptive graph and the invariance representation, it would be possible to visualise how the lack of scalar invariance can influence the differences between countries.

To simplify the comparison, a choice has been made to analyse a sub-group of countries. Belgium has been chosen as the reference country and will be compared with four countries that stood out previously: Switzerland, Norway, Italy and the Netherlands. The graph below will only show the results for those five countries.

On the descriptive graph, it appears that Switzerland and Norway have a higher average than Belgium on both dimensions. When looking at the invariance graph (figure 6.2), it shows that the intercept of Norway and Switzerland is higher in all four variables and in the three rounds than Belgium. This can indicate that a part of the difference in means between the countries is due to a bias in the origin of the scale between Belgium, Switzerland and Norway. Taking into account both graphs, there could still be a true difference between the countries but the difference would probably be smaller than the one represented on the descriptive graph as a part of the difference may be explained by the lack of invariance. For example, the average of the difference in intercepts between Belgium and Switzerland for the two indicators is equal to 0.5 and the difference between the two means for round 8 is 1.15.

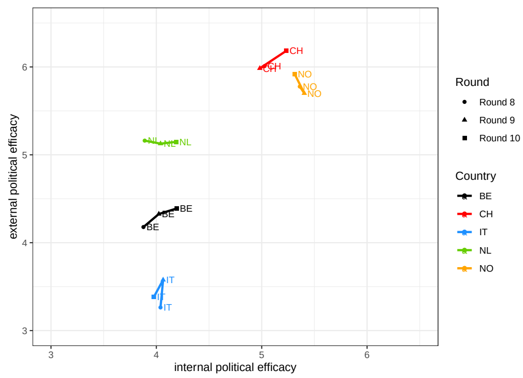


Figure 6.1: Means

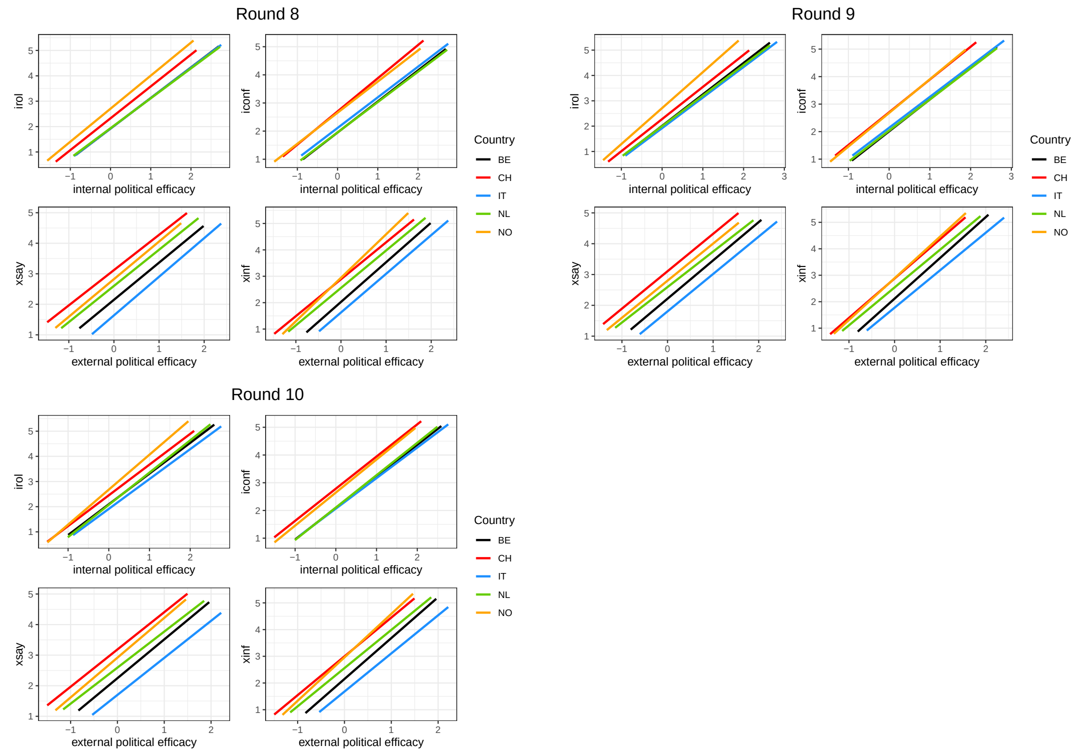


Figure 6.2: Invariance representation

Looking at the Netherlands and Italy in comparison to Belgium, on the internal dimension of political efficacy, the difference in intercepts is very small for both items. Thus, this could mean that the difference on the descriptive graph should not be very biased. When looking at the differences on the internal dimensions, the three countries seem to have similar levels of internal political efficacy.

In contrast, when looking at the external dimensions, the intercepts have a bigger difference. The Netherlands has a higher intercept than Belgium and Italy a smaller intercept than Belgium. On the descriptive graph, a big difference exists in terms of external political efficacy between the three countries and with the difference in intercepts, this difference should be nuanced as some of it may be due to a certain bias.

6.5 Poland and Austria lack of invariance

In this section, the lack of invariance for Poland and Austria is represented in figure 6.3 like in the previous section for the cross-national invariance.

Starting with Poland, it appears, that for the internal political efficacy, the invariance is very small between the three rounds. However, for the external political efficacy, it appears that round 10 has a larger intercept than the other two rounds on both indicators.

Regarding Austria, it appears that only in one indicator (iconf) a difference exists between the three rounds and that the difference between rounds 8 and 10 is slightly larger than the the difference between rounds 8 and 9.

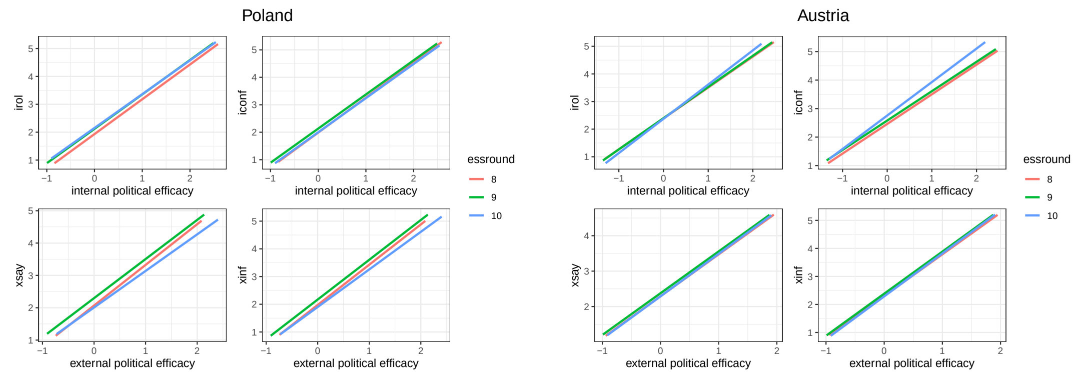


Figure 6.3: Invariance representation

6.6 Latent means

The analysis undertaken in this paper had the objective to check the validity of the direct comparison of political efficacy across different countries in the ESS (see figure 5.1). The previous sections of this chapter have shed light on the fact that there is a lack of invariance between the countries and that when comparing the difference across countries, it should be noted that a part of the difference is explained by the lack of invariance.

The previous section has analysed in parallel the difference between means and the representation of invariance. However, the difference means graph (figure 5.1) is a gross estimation of the latent concept political efficacy as it takes the average of the sum of the two estimators for each dimension. The CFA modelling of the data with its parameter, leads to an estimation of the latent construct where each factor has not necessarily the same influence on each indicator (different loadings and intercepts/thresholds).

It is interesting to see if the CFA model estimation of latent means would demonstrate similar patterns in terms of differences between countries. With the model imposing scalar invariance, a graph with the latent means has been generated for both estimation methods. The graph has been created by running one CFA considering both the country and round as grouping variable. The analysis only took into account the 14 countries that can be modelled through the categorical estimation. Three graphs have been produced below. The first one corresponds to the descriptive graph (figure 6.4). The next graph plots the latent means obtained through the CFA continuous estimation model (figure 6.5). The last graph represents the latent means computed through categorical estimation of the CFA model (figure 6.6).

The first observation that can be made between the three graphs is that the same trend is present in all three of them. Indeed, Switzerland and Norway are the countries with the highest level of political efficacy. Italy is characterised by the smallest level of external political efficacy. Finland and Great Britain have a level of political efficacy that is higher than Belgium but smaller than Norway and Switzerland. Finally, Portugal and Hungary have a smaller level of political efficacy than Belgium.

A second observation is that for Sweden and Slovenia, between the descriptive graph and the two latent means graphs, a difference exists for the estimation of some rounds. In figure 6.4 Sweden has higher average than Belgium for every round whereas in figures 6.5 & 6.6 Sweden (round 10) has a lower average than Belgium. Slovenia has on figure 6.4 has a lower average than Belgium for all three rounds whereas in figures 6.5 & 6.6 Slovenia (round 10) has a higher average than Belgium.

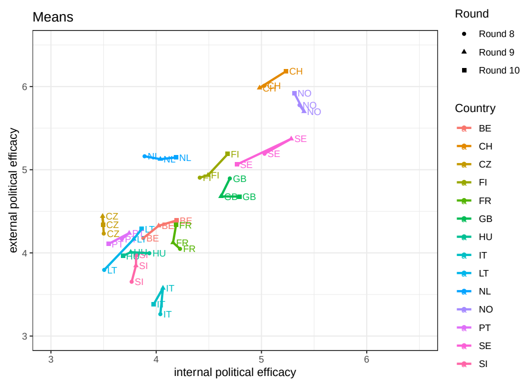


Figure 6.4: Descriptive graph

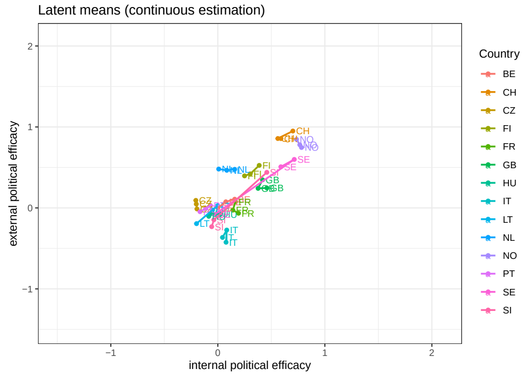


Figure 6.5: Latent means (continuous)

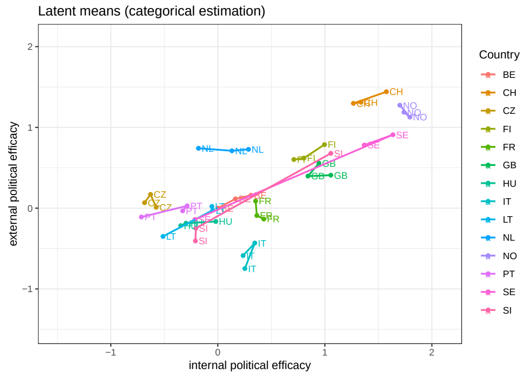


Figure 6.6: Latent means (categorical)

This comparison between the three figures has shed light on the fact that when computing the latent means with the CFA, differences between countries are in most cases consistent with the gross estimation of political efficacy. However, in some cases, the CFA produces different estimations from the descriptive graph.

A few papers about measurement invariance analysis such as Scotto et al. (2021), have produced CFA analyses to test the measurement invariance of a questionnaire. Some of those papers could not reject scalar invariance and then concluded on the differences observed in their descriptive statistics. This section has demonstrated that descriptive statistics do not necessarily give similar results as the CFA analysis. This means that if the CFA analysis does not provide evidence to reject invariance, researchers should also take into account the estimated latent means of the model and not only the descriptive results to make their comparisons.

6.7 Exploration of lack of invariance in political efficacy

The lack of invariance of the total group of countries does not mean that it does not exist sub-groups of scalar invariant countries in terms of political efficacy. Indeed, Xena (2015) did show that even if the analysis taking into account all the countries of the ESS did not reach scalar invariance, some sub-group of countries did reach scalar invariance for political efficacy. This leads us to questioning if the lack of invariance would be caused by another variable that would explain a difference in the citizens' evaluation of the political efficacy. Indeed, in the literature, it has been shown that election rules as well as corruption can influence the level of political efficacy. An exploration based on different variables is proposed to see if an intuition could propose a future track of research to find groups of invariant countries in terms of political efficacy.

Starting with the election rules, there are three possibilities: majoritarian (a party receives a number of seats that is proportional to the percentage of votes cast in its favour in an electoral circumscription), proportional (party with the majority of the vote in one electoral circumscription gets all the seats) or mixed rules (mixes elements of majoritarian and proportional system) (Karp & Banducci, 2008). Figure 6.7 shows in different colours the type of election rules for every country[1]. Based on this graph, there is nothing that really stands out. Indeed, countries with both mixed and proportional rules have both high and low levels of political efficacy. In terms of majoritarian rules, it is difficult to say anything as only two countries follow those elections rules France and Great-Britain.

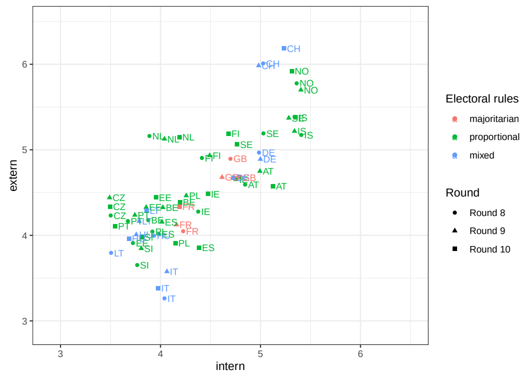


Figure 6.7: Political efficacy & Electoral rules

[1]The information about the electoral rules of the different countries comes from the Electoral Assistance website (*ElecData, Compendium of Electoral Data*, n.d.)

Next, it is known that corruption can influence political efficacy. A transparency index exists to measure how transparent the political system is. The European Research Centre for Anti-Corruption and State-Building has created a government transparency index going from 0 to 20 (with 20 being very transparent, and 0 being not transparent at all) (*Government Transparency Index*, n.d.). On the figure 6.8, transparency does not seem to be linked to political efficacy.

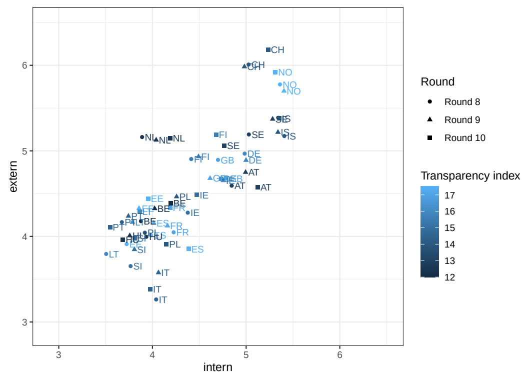


Figure 6.8: Political efficacy & Transparency

In the European Union, the development of democracy did not happen at the same time in every country. It would be interesting to see if the age of the democracy would draw a trend on the graph. Information about democracy age has been compiled by three researchers: Carles Boix, Michael K. Miller, and Sebastian Rosato (2013)[2]. On figure 6.9, there is no clear cut that could explain high or low level of efficacy. However, one thing that appears is that the countries with the lowest levels of efficacy are all quite new democracies.

[2]The data comes from OurWorldData.org: "Data Page: Age of democracy", part of the following publication: Bastian Herre, Lucas Rodés-Guirao, Esteban Ortiz-Ospina and Max Roser (2013) - "Democracy". Data adapted from Boix-Miller-Rosato. Retrieved from https://ourworldindata.org/grapher/age-of-democracy-bmr [online resource]

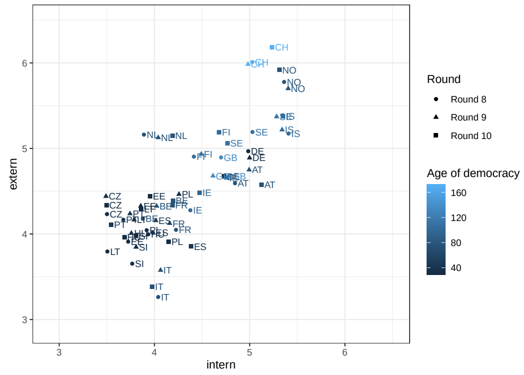


Figure 6.9: Political efficacy & Age of democracy

Finally, the last element investigated is the performance of democracy in the country. Indeed, both dimensions of political efficacy are linked to democracy. For example, the question regarding the ability to participate in a political group also demands that any citizens would have access to political groups in the countries and for example the influence of citizens on government also needs a system where the government is supposed to represent the people who elected them. A democracy index (Economist Intelligence Unit, 2023) has been developed by the Economist. Using the data from the 2022 democratic index, the index goes from 1 to 10 with 10 considered as a high functioning democracy and 1 a low functioning democracy. It has been checked that this index is not a construct based on indicators of political efficacy. Looking at figure 6.10, it appears that the level of democracy influenced the level of political efficacy. Indeed, countries with high index of democracy are countries that have a high level of both internal and external level of political efficacy.

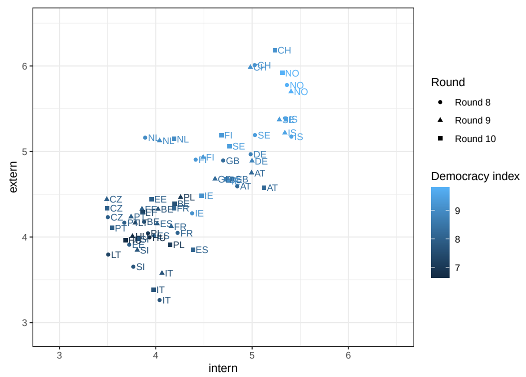


Figure 6.10: Political efficacy & Democracy index

Finally, Xena (2015) demonstrated that scalar invariance could not be accepted across Europe during the first round. Despite having modified the ESS questionnaire since the first round, scalar invariance was still not accepted for a majority of European countries taking part in the ESS in the 8^{th}, 9^{th} and 10^{th} rounds. This brings to light that the ESS questionnaire may need to rethink again the items of political efficacy to have a measurement invariance scale. Furthermore, the ESS should also think to add more than four items in the scale as it was one major limit in this analysis. For a number of countries, the measurement analysis could not be applied because the model was empirically under-identified. Adding new items would reduce the risk of obtaining empirical under-identified models (Roos & Bauldry, 2022).

Chapter 7

Conclusion

Political efficacy is a latent concept that measures both the citizens' own ability to take part in politics as well as the government response to citizens' demands. This concept is divided into two dimensions. The first one is internal political efficacy defined as the evaluation of citizens' capacity to participate in the political system. This means that the citizen has knowledge, confidence and belief about the political system and the role he can play in it. The second dimension is the external political efficacy. Here, the dimension measures the citizens' evaluation of how well the political system responds to them. This studies how the government listens to demands of citizens as well as the power of citizens to influence political elites in power. (Niemi et al., 1991; Bene, 2020)

In the political science research field, it has been demonstrated that political efficacy can influence political participation (Pollock, 1983). In today's society, a trend observed in different countries is that the political participation has been decreasing. In this context, it is important to develop tools to measure accurately political efficacy in different cultures.

The European Social Survey is a well-known cross-national survey that has been developed with an objective to produce high quality data for comparison across time and countries. This survey contains a set of items that measure political efficacy. However, the only measurement invariance analysis produced on the political efficacy scale in the ESS has been done on the first round (Xena, 2015). Since the first round, where scalar invariance was not achieved, the questionnaire has evolved and new items have been proposed to measure political efficacy.

These observations have led to the research question tackled in this paper. *Which level of measurement invariance is reached by the concept of political efficacy in the European Social Survey in the rounds 8, 9 and 10 ?* To answer this question, different multi-group confirmatory factorial analyses have been conducted. All the analyses have been done a first time with the maximum likelihood estimator with the Satorra-Bentler correction and a second time with the robust diagonal weighted least square estimator.

The longitudinal measurement invariance has been investigated. The two estimation methods give similar outcomes. The results are positive for the ESS organisation as only Austria and Poland did not reach scalar invariance but they did achieve metric invariance. The findings of the longitudinal analysis show that the ESS collects data that can be compared across time for the political efficacy concept within a country.

The results of the cross-national invariance analysis are a little less positive for the ESS. Indeed, no round is able to achieve scalar invariance. This means that the ESS political efficacy scale should not be used to compare directly latent means between countries. Furthermore, what appears when the invariance is represented graphically is that for both items of each dimension the difference in intercepts follows the same trend and that the difference between countries is not very big. This does lead to the question that even if the scale is biased, there could still be differences between countries that are not an entirely artificial difference coming from a biased instrument.

In this work, a major limitation in the analysis is the presence of only four items to measure political efficacy in the ESS. If the ESS develops new items for political efficacy, it would be interesting to create a scale with more items as to consolidate findings about measurement invariance of their questionnaire.

Finally, political efficacy has been studied in European countries. The idea to compare this concept in different countries and that the measurement instrument would be invariant probably comes from the impression that as all the countries practice today liberal democracy, the comparison would be made in similar contexts in regards to democracy. The concept of political efficacy does measure how the elected people represent the wishes of the citizens as well as how well the political system is known for citizens to feel competent in it to be able to participate. However, the development of democracy in those different countries happened at different times and in different conditions which means that systems are characterised by different levels of democratic institutions. The difference of democratic development of a country could be the explanation behind the lack of invariance in the different countries. Indeed, if you compare respondents with similar background in terms of socio-economic characteristics in two different countries, one where the institutions are considered as fully democratic and an other where the institutions are barely considered democratic, the citizens evaluation of external political efficacy will be biased as in one country the system does allow citizens influence and in the other, the system allows limited influence of the citizens, for example, comparing political efficacy of Norway and Hungary in the last 10 years could be biased by the decrease of democratic institutions in Hungary (the European parliament has recognised that Hungary has become an electoral autocratic system). This last reflection leads us to a more general question for the future of the ESS questionnaire. Is the lack of scalar invariance of political efficacy due to the instrument in itself or is it due to the different democratic contexts in Europe that leads any instrument of political efficacy to lack scalar invariance for the concept of political efficacy ?

Appendix

Appendix A

Table A1: Participating countries

	Round 8	Round 9	Round 10
AL		√	
AT	√	√	√
BE	√	√	√
BG		√	√
CH	√	√	√
CY		√	√
CZ	√	√	√
DE	√	√	√
DK		√	
EE	√	√	√
ES	√	√	√
FI	√	√	√
FR	√	√	√
GB	√	√	√
GR			√
HR		√	√
HU	√	√	√
IE	√	√	√
IS	√	√	√
IL	√		√
IT	√	√	√
LT	√	√	√
LV		√	√
ME		√	√
MK			√
NL	√	√	√
NO	√	√	√
PL	√	√	√
PT	√	√	√
RO		√	
RS		√	√
RU	√		
SE	√	√	√
SI	√	√	√
SK			√

Table A2: Response rate

	Round 8	Round 9	Round 10
AL		55.6	
AT	52.5	50.8	33.7
BE	56.8	57.6	39.2
BG		69.4	72.5
CH	52.2	51.8	49.5
CY		53.4	14.7
CZ	68.5	67.4	72.8
DE	30.6	27.6	37.0
DK		48.8	
EE	68.4	62.7	47.2
ES	67.7	53.8	35.5
FI	57.7	51.8	41.1
FR	52.4	48.1	39.6
GB	42.8	41.0	20.88
GR			48.0
HR		43.2	43.1
HU	42.7	40.7	40.4
IE	64.5	62.0	36.3
IS	45.8	40.5	33.6
IL	74.4		32.8
IT	49.7		49.8
LT	64.0	59.2	35.6
LV		38.9	23.3
ME		62.3	58.0
MK			60.3
NL	52.9	49.6	35.7
NO	52.8	43.3	37.9
PL	69.6	60.4	39.2
PT	45.0	34.9	41.7
RO		Not given	
RS		57.9	29.6
RU	63.4		
SE	43.0	39.0	37.9
SI	55.9	64.1	54.7
SK		39.6	44.3

Appendix B

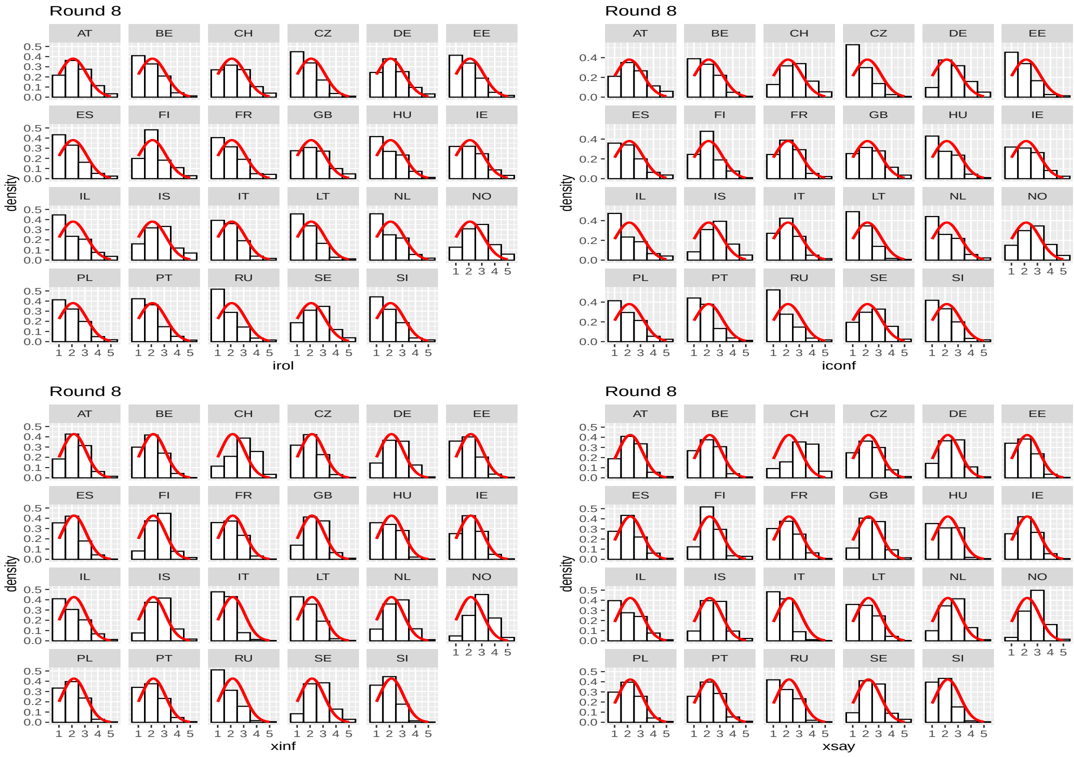


Figure B1: Distribution variables round 8

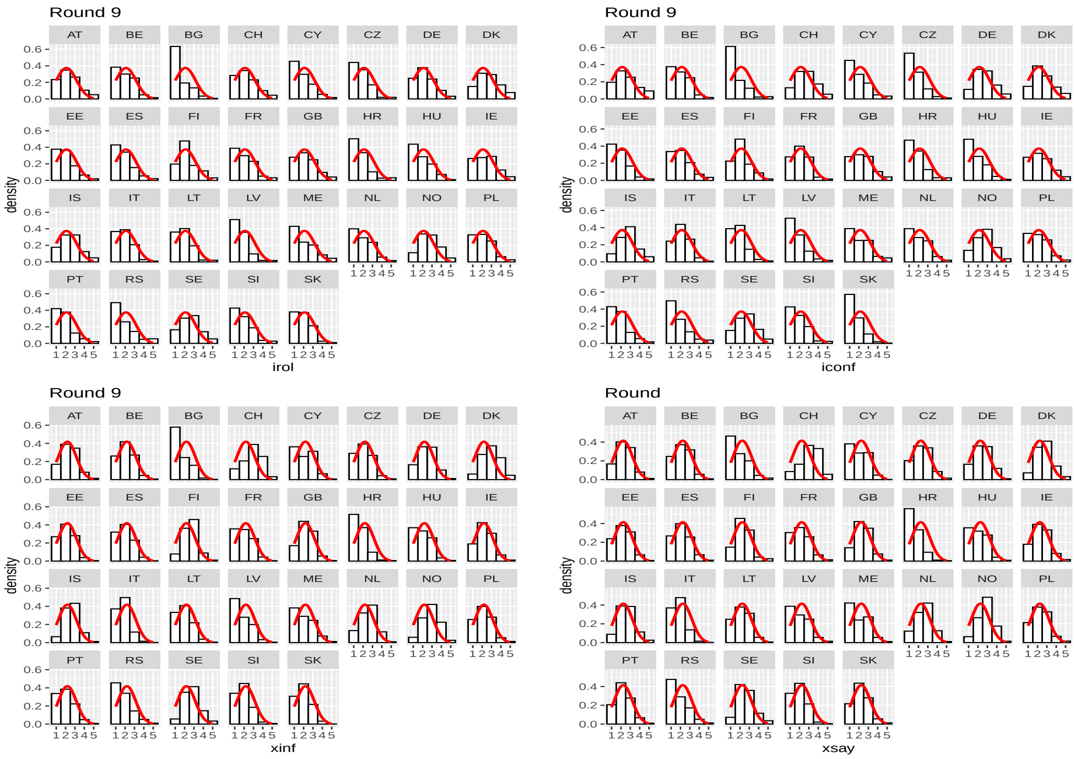


Figure B2: Distribution variables round 9

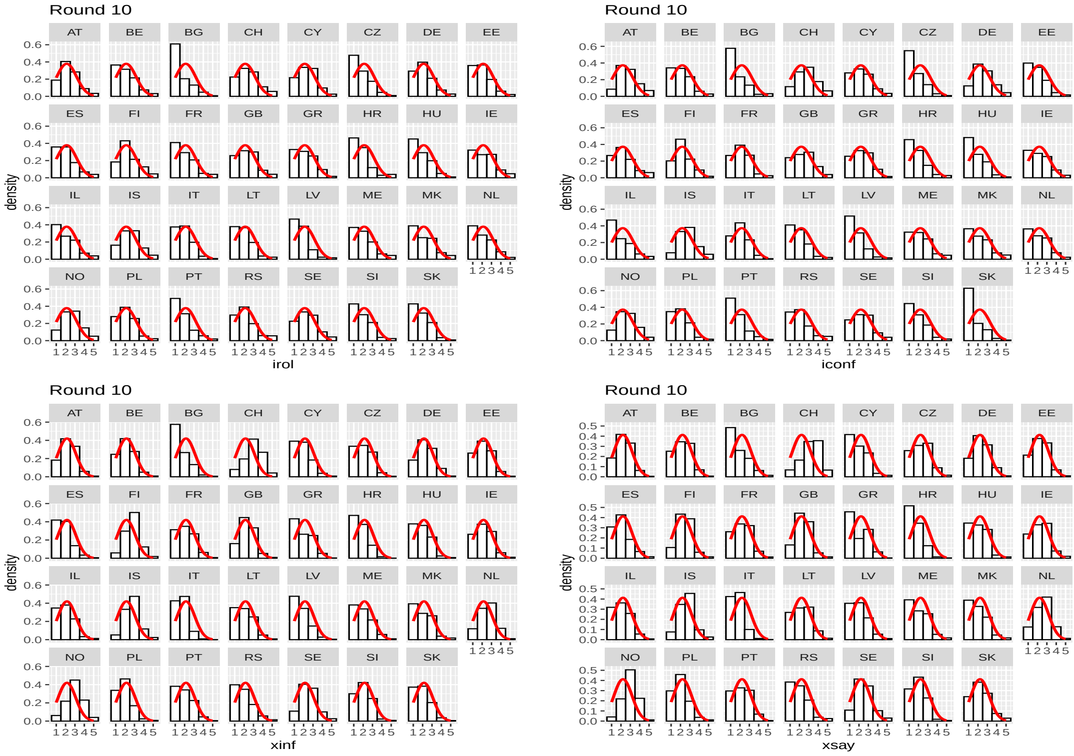


Figure B3: Distribution variables round 10

Table B1: CFI configural model (continuous estimation)

	Round 8	Round 9	Round 10
AT	1.000	0.996	1.000
BE	1.000	1.000	0.997
CH	0.998	0.996	0.999
CZ	0.992	0.998	0.999
EE	1.000	1.000	1.000
FI	0.998	0.991	0.997
FR	0.993	1.000	0.999
GB	1.000	0.998	0.998
HU	1.000	1.000	1.000
IE	1.000	1.000	1.000
IT	0.999	1.000	0.997
LT	0.999	0.984	1.000
NL	0.998	0.999	1.000
NO	1.000	0.999	1.000
PL	0.998	1.000	1.000
PT	1.000	0.997	1.000
SE	1.000	1.000	1.000
SI	1.000	0.996	0.998

Table B2: RMSEA configural model (continuous estimation)

	Round 8	Round 9	Round 10
AT	0.000	0.048	0.000
BE	0.000	0.000	0.037
CH	0.040	0.059	0.025
CZ	0.068	0.029	0.022
EE	0.014	0.000	0.000
FI	0.039	0.075	0.043
FR	0.054	0.000	0.018
GB	0.000	0.033	0.037
HU	0.000	0.000	0.000
IE	0.000	0.000	0.000
IT	0.023	0.002	0.047
LT	0.020	0.079	0.011
NL	0.040	0.024	0.000
NO	0.000	0.022	0.000
PL	0.038	0.000	0.018
PT	0.000	0.035	0.000
SE	0.000	0.000	0.000
SI	0.011	0.060	0.043

Table B3: Longitudinal invariance: CFI (continuous estimation)

	configural	metric	scalar	$\Delta metric$	$\Delta scalar$
AT	0.999	0.999	0.977	0.000	-0.022
BE	0.999	1.000	1.000	0.001	0.000
CH	0.998	0.998	0.997	0.000	-0.001
CZ	0.997	0.996	0.996	-0.001	0.000
EE	1.000	0.999	0.990	-0.001	-0.009
FI	0.996	0.996	0.996	0.000	0.000
FR	0.998	0.999	0.995	0.001	-0.004
GB	0.999	1.000	0.998	0.001	-0.002
HU	1.000	1.000	1.000	0.000	0.000
IE	1.000	1.000	0.995	0.000	-0.005
IT	0.999	0.999	0.997	0.000	-0.002
LT	0.995	0.992	0.987	-0.003	-0.005
NL	0.999	1.000	1.000	0.001	0.000
NO	1.000	1.000	0.998	0.000	-0.002
PL	0.999	0.999	0.985	0.000	-0.014
PT	1.000	1.000	0.999	0.000	-0.001
SE	1.000	1.000	0.998	0.000	-0.002
SI	0.998	0.995	0.990	-0.003	-0.005

Table B4: Longitudinal invariance: RMSEA (continuous estimation)

	configural	metric	scalar	$\Delta metric$	$\Delta scalar$
AT	0.017	0.015	0.056	-0.002	0.041
BE	0.022	0.007	0	-0.015	-0.007
CH	0.043	0.028	0.028	-0.015	0.000
CZ	0.044	0.032	0.025	-0.012	-0.007
EE	0.000	0.019	0.046	0.019	0.027
FI	0.055	0.035	0.028	-0.02	-0.007
FR	0.029	0.015	0.024	-0.014	0.009
GB	0.025	0.004	0.017	-0.021	0.013
HU	0.000	0.000	0.000	0.000	0.000
IE	0.000	0.010	0.029	0.01	0.019
IT	0.030	0.016	0.025	-0.014	0.009
LT	0.052	0.043	0.043	-0.009	0.000
NL	0.026	0.012	0.000	-0.014	-0.012
NO	0.000	0.000	0.016	0.000	0.016
PL	0.024	0.020	0.055	-0.004	0.035
PT	0.001	0.000	0.008	-0.001	0.008
SE	0.000	0.000	0.020	0.000	0.020
SI	0.041	0.039	0.047	-0.002	0.008

Table B5: Poland longitudinal invariance

		CFI		RMSEA
	Δ_metric	Δ_scalar	Δ_metric	Δ_scalar
Round 8-9	0.001	-0.002	-0.013	0.007
Round 9-10	-0.001	-0.011	0.005	0.042
Round 8-10	-0.002	-0.015	0.001	0.036

Table B6: Austria longitudinal invariance

		CFI		RMSEA
	Δ_metric	Δ_scalar	Δ_metric	Δ_scalar
Round 8-9	0.000	-0.006	-0.005	0.019
Round 9-10	0.000	-0.009	-0 .012	0.021
Round 8-10	0.000	-0.035	0 .009	0.064

Table B7: Latent means difference p-values (continuous estimation))

		Intern		Extern
	Round 9	Round 10	Round 9	Round 10
AT	/	/	/	/
BE	0.0131	0.0000	0.0027	0.0002
CH	**0.6238**	0.0070	0.7464	0.0050
CZ	**0.8560**	**0.9095**	0.0003	**0.0549**
EE	0.0232	0.0002	0.0000	0.0000
FI	**0.2026**	0.0000	**0.5568**	0.0000
FR	0.0492	**0.2950**	**0.1889**	0.0000
GB	**0.2316**	**0.6290**	0.0002	0.0056
HU	0.0257	0.0006	**0.7736**	**0.3810**
IE	0.0000	**0.1215**	0.0000	0.0206
IT	**0.7434**	**0.2731**	0.0000	0.0018
LT	0.0001	0.0000	0.0000	0.0000
NL	0.0459	0.0001	**0.7120**	**0.9099**
NO	**0.7275**	**0.4938**	**0.1459**	**0.2284**
PL	/	/	/	/
PT	**0.5144**	**0.1068**	**0.5921**	**0.2822**
SE	0.0022	0.0007	0.0096	0.0048
SI	**0.5166**	**0.4637**	0.0335	0.0000

Table B8: Configural model: CFI (categorical estimation)

	Round 8	Round 9	Round 10
BE	1.000	1.000	0.999
CH	0.999	0.998	1.000
CZ	0.998	0.999	1.000
FI	0.999	0.996	0.998
FR	0.998	1.000	1.000
GB	1.000	0.999	0.999
HU	1.000	1.000	1.000
IT	1.000	1.000	0.999
LT	1.000	0.997	1.000
NL	0.999	1.000	1.000
NO	1.000	0.999	1.000
PT	1.000	0.999	1.000
SE	1.000	1.000	1.000
SI	0.999	0.999	0.999

Table B9: Configural model: RMSEA (categorical estimation)

	Round 8	Round 9	Round 10
BE	0.008	0.000	0.052
CH	0.055	0.071	0.033
CZ	0.084	0.048	0.035
FI	0.059	0.103	0.069
FR	0.058	0..000	0.029
GB	0.000	0.045	0.049
HU	0.028	0.000	0.000
IT	0.034	0.022	0.066
LT	0.038	0.099	0.006
NL	0.050	0.038	0.000
NO	0.000	0.030	0.000
PT	0.011	0.032	0.000
SE	0.000	0.000	0.000
SI	0.045	0.063	0.058

Table B10: Longitudinal invariance: CFI (categorical estimation)

	configural	metric	scalar	Δ_{metric}	Δ_{scalar}
BE	1.000	1.000	0.999	0.000	-0.001
CH	0.999	0.999	0.999	0.000	0.000
CZ	0.999	0.999	0.998	0.000	-0.001
FI	0.998	0.999	0.997	0.001	-0.002
FR	0.999	1.000	0.998	0.001	-0.002
GB	0.999	1.000	1.000	0.001	0.000
HU	1.000	1.000	0.999	0.000	-0.001
IT	1.000	1.000	1.000	0.000	0.000
LT	0.999	0.999	0.999	0.000	0.000
NL	1.000	1.000	1.000	0.000	0.000
NO	1.000	1.000	0.997	0.000	-0.003
PT	1.000	1.000	0.998	0.000	-0.002
SE	1.000	1.000	0.999	0.000	-0.001
SI	0.999	0.999	0.998	0.000	-0.001

Table B11: Longitudinal invariance: RMSEA (categorical estimation)

	configural	metric	scalar	Δ_{metric}	Δ_{scalar}
BE	0.027	0.007	0.019	-0.02	0.012
CH	0.054	0.033	0.017	-0.021	-0.016
CZ	0.059	0.030	0.028	-0.029	-0.002
FI	0.079	0.039	0.031	-0.04	-0.008
FR	0.033	0.015	0.019	-0.018	0.004
GB	0.032	0.000	0.002	-0.032	0.002
HU	0.006	0.023	0.022	0.017	-0.001
IT	0.046	0.018	0.009	-0.028	-0.009
LT	0.061	0.041	0.020	-0.02	-0.021
NL	0.033	0.000	0.000	-0.033	0.000
NO	0.015	0.017	0.022	0.002	0.005
PT	0.014	0.000	0.026	-0.014	0.026
SE	0.000	0.011	0.023	0.011	0.012
SI	0.055	0.035	0.032	-0.02	-0.003

Table B12: Latent means difference p-values (categorical estimation)

		Intern		Extern
	Round 9	Round 10	Round 9	Round 10
BE	0.0435	0.0019	0.0060	0.0010
CH	**0.5278**	0.0063	**0.6710**	0.0119
CZ	**0.7030**	**0.2841**	0.0003	**0.2231**
FI	**0.2440**	0.0001	**0.7408**	0.0000
FR	**0.2978**	**0.3799**	**0.4352**	0.0000
GB	**0.2288**	**0.2931**	0.0002	0.0031
HU	0.0130	0.0020	**0.7330**	**0.3528**
IT	**0.1022**	**0.7274**	0.0000	0.0000
LT	0.0000	0.0001	0.0000	0.0000
NL	0.0062	0.0002	**0.5089**	**0.8090**
NO	**0.5295**	**0.6425**	**0.1433**	**0.1137**
PT	**0.8246**	0.0034	**0.3611**	**0.1862**
SE	0.0036	0.0005	0.0096	0.0264
SI	**0.8852**	**0.9494**	0.0035	0.0000

Bibliography

Balch, G. I. (1974). Multiple indicators in survey research: The concept "sense of political efficacy". *Political Methodology, 1*(2), 1-43.

Bene, M. (2020). Does context matter? A cross-country investigation of the effects of the media context on external and internal political efficacy. *International Journal of Comparative Sociology, 61*(4), 264-286.

Boix, C., Miller, M., & Rosato, S. (2013). A complete data set of political regimes, 1800–2007. *Comparative Political Studies, 46*(12), 1523-1554.

Bollen, K. A. (1989). *Structural equations with latent variables.* Wiley.

Brown T. & Moore M.(2012) Confirmatory Factor Analysis. In: Hoyle, R. (Ed.) *Handbook of structural equation modelling (pp.361-379).* The Guilford Press.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research (2nd ed.).* The Guilford Press.

Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-Cultural Psychology, 34*(2), 155-175.

Campbell, A., Gurin, D. and Miller, W.E. (1954). *The Voter Decides.* Row, Peterson, and Company.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology, 31*(2), 187-212.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233-255.

Chou, C-P. & Bentler, P. (1995). Estimates and Tests in Structural Equation Modelling. In: Hoyle, R (Ed.) *Structural equation modeling: Concepts, issues and applications (pp.37-55).* Sage.

D'Urso, E. D., De Roover, K., Vermunt, J. K., & Tijmstra, J. (2022). Scale length does matter: Recommendations for measurement invariance testing with categorical factor analysis and item response theory approaches. *Behavior Research Methods, 54*(5), 2114-2145.

"Data Page: Age of democracy", part of the following publication: Bastian Herre, Lucas Rodés-Guirao, Esteban Ortiz-Ospina and Max Roser (2013) - "Democracy". Data adapted from Boix-Miller-Rosato. Retrieved from https://ourworldindata.org/grapher/age-of-democracy-bmr [online resource]

Data Quality Assessment | European Social Survey. (n.d.). Retrieved December 2023 from https://www.europeansocialsurvey.org/methodology/ess-methodology/data-quality-assessment

Economist Intelligence Unit (2023), Economist Intelligence Unit (2021) via gapminder – processed by Our World in Data. "Democracy" [dataset]. Economist Intelligence Unit (2023), Economist Intelligence Unit (2021) via gapminder [original data].

ElecData, Compendium of Electoral Data. (n.d.). Electoral Assistance. Retrieved May 23, 2024, from https://www.coe.int/en/web/electoral-assistance/elecdata

ESS (1999).*The European Social Survey (ESS) – a research instrument for the social science in Europe.* Strasbourg: ESF.

ESS Round 8 Question Design Template -New Core Items Political Efficacy. (n.d.). Retrieved May 23, 2024, from https://europeansocialsurvey.org/sites/default/files/2023-06/ESS8_political_efficacy_final_temp.pdf

ESS Rounds 1-5: Concepts, indicators, indices and their quality| European Social Survey. (n.d.). Retrieved May 23, 2024, from https://www.europeansocialsurvey.org/sites/default/file
06/ESS1-5_indices_quality.pdf

European Social Survey European Research Infrastructure (ESS ERIC). (2023c). ESS10 Self-completion - integrated file, edition 3.1 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess10sce03_1

European Social Survey European Research Infrastructure (ESS ERIC). (2023b). ESS9 - integrated file, edition 3.2 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/ess9e03_2

European Social Survey European Research Infrastructure (ESS ERIC). (2023a). ESS8 - integrated file, edition 2.3 [Data set]. Sikt - Norwegian Agency for Shared Services in Education and Research. https://https://doi.org/10.21338/ess8e02_3

European Social Survey European Research Infrastructure (ESS ERIC). (2002). ESS1 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess1-2002

European Social Survey European Research Infrastructure (ESS ERIC). (2021). ESS2 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/NSD-ESS2-2004

European Social Survey European Research Infrastructure (ESS ERIC). (2018a). ESS3 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/NSD-ESS3-2006

European Social Survey European Research Infrastructure (ESS ERIC). (2008). ESS4 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess4-2008

European Social Survey European Research Infrastructure (ESS ERIC). (2010). ESS5 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess5-2010

European Social Survey European Research Infrastructure (ESS ERIC). (2012). ESS6 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess6-2012

European Social Survey European Research Infrastructure (ESS ERIC). (2018b). ESS7 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/NSD-ESS7-2014

European Social Survey European Research Infrastructure (ESS ERIC). (2016). ESS8 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess8-2016

European Social Survey European Research Infrastructure (ESS ERIC). (2018c). ESS9 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/nsd-ess9-2018

European Social Survey European Research Infrastructure (ESS ERIC). (2022). ESS10 Data Documentation. Sikt - Norwegian Agency for Shared Services in Education and Research. https://doi.org/10.21338/NSD-ESS10-2020

Finney, S. J., & DiStefano, C. (2006). Nonnormal and Categorical Data in Structural Equation Models. In: G. R. Hancock, & R. O. Mueller (Eds.), *A Second Course in Structural Equation Modeling (pp. 269-314).* Information Age Publishing.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491.

Fraile, M., & de Miguel Moyer, C. (2022). Risk and the gender gap in internal political efficacy in europe. *West European Politics, 45*(7), 1462-1480.

Government Transparency Index. (n.d.). ERCAS. https://www.againstcorruption.eu/ercas-projects/transparencyindex/

He, J., & van de Vijver, F. (2012). Bias and Equivalence in Cross-Cultural Research. *Online Readings in Psychology and Culture, 2*(2), 3-19.

Herk, H. V., & Goldman, S. P. K. (2022). The Advancement of Measurement Invariance Testing in Cross-Cultural Research in the Period 1999–2020. Executing Rather Than Scrutinizing? In H. Baumgartner, & B. Weijters (Eds.), *Measurement in Marketing* (pp. 95-119). (Review of Marketing Research; Vol. 19). Emerald Publishing.

History | European Social Survey. (n. d.). Retrieved April 2024 from on https://www.europeansocialsurvey.org/about/history

Horn, J. L., & McArdle, J. J. (1992). A Practical and Theoretical Guide to Measurement Invariance in Aging Research. *Experimental Aging Research, 18*(3), 117–144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Husson, F., Lê, S., & Pagès, J. (2017). *Exploratory multivariate analysis by example using R (2nd ed.).* CRC.

Items dropped from the core questionnaire in ESS Rounds 5 and 6 | European Social Survey. (n.d.). Retrieved May 23, 2024, from https://www.europeansocialsurvey.org/sites/default/files 06/ESS5-6_core_changes.pdf

Johnson, T. (1998). Approaches to Equivalence in Cross-Cultural and Cross-National Surveys. *ZUMA Nachrichten Spezial: Cross-Cultural Survey Equivalence, 3*, 1–40.

Kankaraš, M., & Moors, G. (2010). Researching measurement equivalence in cross-cultural studies. *Psihologija, 43*(2), 121-136.

Karp, J. A., & Banducci, S. A. (2008). Political Efficacy and Participation in Twenty-Seven Democracies: How Electoral Systems Shape Political Behaviour. *British Journal of Political Science, 38*(2), 311–334.

Kline, R. B. (2016). *Principles and practice of structural equation modeling (4th ed.).* Guilford press.

Lane, R. (1959). *Political life. Why people get involved in.* The Free Press.

Lebart, L., Morineau, A., & Piron, M. (2000). *Statistique exploratoire multidimensionnelle (3e ed.).* Dunod.

LeDuc, L. (1976). Measuring the sense of political efficacy in canada: Problems of measurement equivalence. *Comparative Political Studies, 8*(4), 490-500.

Lei, P-W. & Wu, Q.(2012) Estimation in Structural Equation Modeling. In: Hoyle, R. (Ed.) *Handbook of structural equation modelling (pp.164-180).* The Guilford Press.

Li, C. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949.

Mokken, R. J. (1969). Dutch-American comparisons of the 'sense of political efficacy': some remarks on cross-cultural 'robustness'. *Acta Politica, 4: 1968/1969*(4), 425-448.

Morrell, M. E. (2003). Survey and experimental evidence for a reliable and valid measure of internal political efficacy. *Public Opinion Quarterly, 67*(4), 589-602.

Muthén, B (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika, 49*, 115–132.

Niemi, R. G., Craig, S. C., & Mattei, F. (1991). Measuring internal political efficacy in the 1988 national election study. *The American Political Science Review, 85*(4), 1407-1413.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*, 443-460.

Pollock, P. H. (1983). The participatory consequences of internal and external political efficacy. *The Western Political Quarterly, 36*(3), 400-409.

Reichert, F. (2016). How internal political efficacy translates political knowledge into political participation: Evidence from Germany. *Europe's Journal of Psychology, 12*(2), 221-241.

Roos, J. M., & Bauldry, S. (2022). *Confirmatory factor analysis.* Sage.

Scotto, T. J., Xena, C., & Reifler, J. (2021). Alternative measures of political efficacy: The quest for cross-cultural invariance with ordinally scaled survey items. *Frontiers in Political Science, 3*, 1-14.

Shore, J., Rapp, C., & Stockemer, D. (2019). Health and political efficacy in context: What is the role of the welfare state? *International Journal of Comparative Sociology, 60*(6), 435-457.

Singh, J. (1995). Measurement issues in cross-national research. *Journal of International Business Studies, 26*(3), 597-619.

Solijonov, A. (2016). Voter Turnout Trends around the World. *International Institute for Democracy and Electoral Assistance.* 1-54.

Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in Cross-National consumer research. *The Journal of Consumer Research, 25*(1), 78-107.

Sulitzeanu-Kenan, R., & Halperin, E. (2013). Making a difference: Political efficacy and policy preference construction. *British Journal of Political Science, 43*(2), 295-322.

van de Vijver, F. J. R., & Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research.* Cambridge University Press.

van de Vijver, Fons J. R., & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment : Official Organ of the European Association of Psychological Assessment, 13*(1), 29-37.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4-70.

Vecchione, M., Caprara, G. V., Caprara, M. G., Alessandri, G., Tabernero, C., & González-Castro, J. L. (2014). The perceived political self-efficacy Scale–Short form (PPSE-S): A validation study in three mediterranean countries. *Cross-Cultural Research, 48*(4), 368-384.

Welkenhuysen-Gybels, J., van de Vijver, F. J. R., & Cambré, B. (2007). A comparison of method for the evaluation of construct equivalence in a multigroup setting. In: G. Loosveldt, M. Swyngedouw, & B. Cambré (Eds.), *Measuring meaningful data in social research (pp. 357-371).* Acco.

Widaman, K.F., & Reise, S.P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In: KJ. Bryant, ME. Windle, SG. West (Eds.), *The Science of Prevention: Methodological Advances from Alcohol and Substance Abuse Research (pp. 281-324).* American Psychological Association.

Xena, C. (2015). *Old Measures, New Implications: The Meaning of Political Efficacy across Political Contexts.* Colchester: University of Essex. PhD thesis.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté des sciences
Place des Sciences, 2 bte L6.06.01, 1348 Louvain-la-Neuve, Belgique | www.uclouvain.be/sc