

OECD Working Papers on Public Governance No. 38

Testing the evidence, how good are public sector responsiveness measures and how to improve them?

Santiago González

https://dx.doi.org/10.1787/c1b10334-en

OECD

Organisation for Economic Co-operation and Development

GOV/PGC(2020)11

For Official Use

English - Or. English

7 May 2020

PUBLIC GOVERNANCE DIRECTORATE
PUBLIC GOVERNANCE COMMITTEE

Testing the evidence, how good are public sector responsiveness measures and how to improve them?

Keywords: responsiveness, political efficacy, satisfaction with services, accuracy, reliability, validity, household surveys, government, public management, public governance indicators
JEL Classification: A13, C46, C83, H4, H11, H83 I1,I2, I38

santiago.gonzalez@oecd.org

JT03461517

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Abstract

This paper analyses two common uses of the responsiveness concept in the public management and political science literature: external political efficacy and satisfaction with health and education services. The decline of people sense of influence in public affairs and perceptions about the quality of public services are two key concerns affecting policymaking. The fact that responsiveness measures are increasingly being collected in non-official and official household surveys and the range of covariates available make it possible to test their statistical accuracy. Accuracy encompasses both reliability (i.e. if the measure produces consistent information over time) and validity (i.e. if the measure reflects the underlying concept being measured). This paper finds good evidence on the accuracy of political efficacy measures. Although no sufficiently strong evidence on the accuracy of satisfaction with health metrics is stronger than for education services signaling the relevance of other aspects such as direct exposure to the service and its intensity, as well as the different attributes shaping satisfaction levels. Findings from this paper support some of the conclusions in the Responsiveness chapter of the UN Citi Praia Handbook on Governance Statistics.

Acknowledgments

The research conducted for this paper has informed the Responsiveness chapter of the UN Citi Praia Group Handbook on Governance Statistics. Invaluable comments have been provided by Monica Brezzi (OECD Public Governance Directorate), Marco Mira d'Ercole (OECD Statistics and Data Directorate, Andrew Blazey (OECD Public Governance Directorate), Zsuzsanna Lonti (OECD Public Governance Directorate). Comments to previous versions of this paper were provided by Oscar Jaimes Bello from the Mexican National Institute of Statistics and Geography (INEGI), Leonardo Athias from the Brazilian Institute of Geography and Statistics (IBGE), Mehdi Hussain from the Department for International Development (DFID) of the United Kingdom, Alexandra Wilde from the United Nations Development Programme (UNDP) Oslo Governance Centre, Patrick Spearing from the United Nations Department of Economic and Social Affairs (UNDESA) and Ionica Berevoescu from the United Nations Entity for Gender and Equality (UN Women).

Acronyms

AGSS	Australian General Social Survey
ANES	American National Election Studies
CPS	Centre for Political Studies at the University of Michigan
CSES	The Comparative Study of Electoral Systems
ENCUP	Mexican National Survey on Political Culture and Citizenship
EQI	The Quality of Government Index Survey
EQLS	European Quality of Life Survey
ESS	European Social Survey
GWP	Gallup World Poll
IAEG-SDGs	Inter-Agency and Expert group on SDGs indicators
IAT	Implicit Association Test
IDEA	Institute for Democracy and Electoral Assistance
INEGI	Mexican National Institute of Statistics and Geography
ISSP	International Social Survey Programme
NSOs	National Statistical Offices
PaRIS	OECD Patient-Reported Indicator Survey
PIAAC	OECD Survey of Adult Skills
PIRLS	Progress in the International Literacy Study
PISA	OECD Programme for International Student Assessments
PREMS	OECD Patient-Reported Experience Measures

PROMS	OECD Patient Reported Outcome Measures
PSC	Political Support in Canada experiment
SDGs	Sustainable Development Goals
SHaSA	Strategy for the Harmonisation of Statistics in Africa
SRC	Survey Research Centre at the University of Michigan
UNDP	United Nations Development Programme
UNSC	United Nations Statistical Commission

Table of contents

Abstract 2

Acknowledgments 3

Acronyms 4

Executive Summary 8

1 Introduction 9

2 Importance of the dimension and available statistics 11
- 2.1 Political Efficacy: importance of the dimension 11
- 2.2 Political Efficacy: available statistics 11
- 2.3 Satisfaction with education and healthcare services: importance of the dimension 15
- 2.4 Satisfaction with education and healthcare services: available statistics 16

3 Accuracy of responsiveness measures 19
- 3.1 Reliability of political efficacy measures 19
- 3.2 Validity of political efficacy measures 22
- 3.3 Reliability of satisfaction with education and healthcare measures 28
- 3.4 Validity of satisfaction with education and healthcare measures 31

4 Conclusions and statistical agenda ahead 38
- References 39

Annex A. Table external political efficacy questions in selected survey modules. 44

Annex B. Questions on satisfaction with health and education services in selected survey modules. 47

Annex C. Results of the confirmatory factor analysis political efficacy modules 50

Tables

Table 2.1. Household surveys including political efficacy questions 15
Table 2.2. Surveys including satisfaction with services questions 18
Table 3.1. Intra-wave correlation between measures of political efficacy in waves 7, 8 and 9 of the European Social Survey 21

Table 3.2. Intrawave correlation between waves 3, 4 and 5 of the International Social Survey Programme (ISSP) 22
Table 3.3. Pairwise correlations between measures of political efficacy and other governance variables 25
Table 3.4. Cross-country pairwise correlation between measures of political efficacy and proxy concepts 27
Table 3.5. Summary evaluation of the accuracy of survey measures of system responsiveness 28
Table 3.6. Cross-country pairwise correlation between measures of satisfaction with health care services from different household surveys, 2006-2017 28
Table 3.7. Cross-country pairwise correlation between measures of satisfaction with education services from different household surveys, 2006-2017 29
Table 3.8. Intrawave correlation of measures of satisfaction with health services between waves 1 to 8 of the European Social Survey 29
Table 3.9. Intrawave correlation of measures of satisfaction with education services between waves 1 to 8 of the European Social Survey 29
Table 3.10. Intrawave correlation of measures of satisfaction with health services between different waves of Latinbarometer 30
Table 3.11. Intrawave correlation of measures of satisfaction with education services between different waves of Latinbarometer 30
Table 3.12. Intrawave correlation waves of measures of satisfaction with health care services between different waves of the European Quality of Life Survey 30
Table 3.13. Intrawave correlation of measures of satisfaction with education services between waves of the European Quality of Life Survey 31
Table 3.14. Cross country correlation between measures of users' satisfaction with health care services and measures of different obstacles in accessing them, 2016 36
Table 3.15. Summary evaluation of the accuracy of survey measures of satisfaction with health and education service 37

Figures

Figure 2.1. Having a say in what the government does in the US 12
Figure 2.2. Having a say on what the government does 13
Figure 2.3. Having a say in what government does, 2014, 2016 AND 2018 14
Figure 2.4. Satisfaction with health and education in OECD member countries, 2019 17
Figure 2.5. Satisfaction with health and education in Latin America, 2016 17
Figure 3.1. Cross-country correlation of political efficacy as measured in PIAAC and ESS, around 2014 20
Figure 3.2. Cross-country correlation between measures of political efficacy in ISSP and ESS, 2016 21
Figure 3.3. Item-specific non-response rates in the European Social Survey, 2014, 2016 and 2018 23
Figure 3.4. Item specific non-response rates for questions on political efficacy before and after the introduction of a neutral category of response in ENAS 24
Figure 3.5. Political efficacy and trust in government in the United States, 1958-2016 26
Figure 3.6. Political efficacy and voter turnout in the United States, 1964-2016 26
Figure 3.7. Item specific non-response in the Gallup Wold Poll 32
Figure 3.8. Item-specific non-response in the EQLS, 2016 32
Figure 3.9. Cross-country correlations between measures of satisfaction with government and measures of satisfaction with health care and education services 33
Figure 3.10. Cross-country correlations between measures of trust in government and measures of satisfaction with health care and education services 34
Figure 3.11. Users' satisfaction with a general practitioner, family doctor or health centre services by experience, 2016 35
Figure 3.12. Satisfaction with hospital or medical specialist services by experience, 2016 36

Executive Summary

This paper analyses the statistical accuracy of some available metrics to measure public sector responsiveness. Two measures are investigated in the paper: the perception of having a say in what the government does (known as external political efficacy) and satisfaction with education and healthcare services.

These two measures are recognised as key outcomes of public administration in OECD Government at a Glance and underpin the monitoring of some SDG 16 indicators, such as *proportion of the population satisfied with their last experience of public services* and *proportion of the population who believe decision making is inclusive and responsive*, that countries around the world are expected to measure in the near future.

In order to test the statistical accuracy of public sector responsiveness measures this paper tests systematically their reliability and validity, using a range of available household surveys carried out at national and international levels by National Statistical Offices and non-official data providers. Reliability is the degree to which a measure produces consistent information over time and across different measurement vehicles. In turn, validity assesses the extent to which the measure actually captures the underlying concept to be measured.

The paper finds that political efficacy measures perform well enough in terms of reliability and validity for data providers to collect these data based on their own surveys. Similarly, there is good evidence of reliability and one of the three considered aspects of validity for measures of satisfaction with healthcare services, In turn, the accuracy of satisfaction with education measures is weaker for all reliability and validity criteria assessed in this paper.

Finally, this paper suggests a number of further statistical tests for improving public sector responsiveness measures. In the case of political efficacy it would be important to test whether the selection of different wording (e.g. political system and government) influences responses to surveys. It will also be desirable to agree on a common response scale or a system of equivalences between the different scales currently being implemented in different surveys. Testing how political efficacy questions are influenced by the closeness of elections and the availability of new technologies facilitating participation and engagement could provide additional valuable information.

The evidence presented here supports the view that biases towards the government or the public sector influence negatively responses on satisfaction with health and education services. Further investigating the size of these biases and if they are mitigated by more precise questions about service attributes (e.g. access, quality) could be useful for refining existing satisfaction measures. Additionally, there is evidence that recent experiences with health services influence reported satisfaction. However, it could be possible to investigate the effects of the intensity (i.e. a more distant or closer interaction) of experience in levels of satisfaction with services. Further work could also advance in standardizing the measurement of different attributes shaping satisfaction (e.g. financial and geographic access, quality of services) in household surveys for each of the services.

1 Introduction

Policy makers are increasingly confronted with the need of high quality indicators to inform their decisions and substantiate their policy choices in different dimensions of public governance. While the amount of evidence available has steadily grown over the past years, little is known about the statistical quality of these indicators. In a similar spirit to the OECD *Guidelines on Measuring Trust*, this paper addresses this gap by systematically assessing the statistical accuracy of responsiveness measures and identifying areas where further improvement is possible. It is a methodological piece intended for National Statistical Offices (NSOs) and other data producers, as well for policy makers interested in improving the quality of public governance and particularly responsiveness metrics.

Responsiveness has been defined as "how quickly and well a person or organization reacts to something"[1]. One insight from this definition is that responsiveness implies a challenge or *stimuli* that triggers a reaction. The term has been used in several contexts as, for example, the effects of mothers' *response* to the frequency and duration of infant crying (Bell and Salter Ainsworth, 1972), the *reaction* of the human immune system to a specific substance (Piani et al, 2000) or the motivations that will make companies *adopt* an environmentally friendly approach (Bansal and Roth, 2017).

Political science and public management have analysed the responsiveness of public institutions. Yet these institutions are different in nature; some authors focus on institutions in the political sphere and emphasize the role of political competition and how the degree of contestation influences policy responsiveness (Hobolt and Klemmensen, 2008). In turn, others have emphasized the institutional aspects of responsiveness as shaped by people expectations and interactions with public institutions (Ivanyna and Shah, 2018). Common to most uses in fields related to public governance is that institutional responsiveness is driven by people's preferences and that public institutions react to those preferences.

This paper looks at the concept of external political efficacy, also called system responsiveness, and its measurement, as well as one specific aspect of institutional responsiveness, in the field of satisfaction with selected public services (i.e. health and education). While not comprehensive[2] the selection of these two dimensions is not arbitrary. To begin with, the decline of people's sense of influence in public affairs through traditional channels and widespread perceptions about deteriorating quality of public services are two key concerns of policymaking. It is then important to have solid measures of these phenomena to better understand their determinants and address their deterioration.

Moreover, both political efficacy and satisfaction with public services are recognised as key outcomes of public administrations in the OECD flagship report Government at a Glance (OECD 2019a; OECD 2017a; OECD 2017b). Survey measures of political efficacy have been included as a headline indicator of the OECD Well-Being framework (OECD 2017a; OECD 2020).

These two dimensions of responsiveness are also topical for data producers as evidenced by the emphasis placed on public services and measuring different aspects of their access and quality in the fourth and latest available wave of the European Quality of Life Survey released in 2016 (Eurofund, 2019) and by the inclusion of political efficacy questions in the source questionnaire of the European Social Survey (ESS) since wave eight, fielded in 2016.

The aspects of responsiveness treated in this paper are also targets of the Inter Agency Expert Group (IAEG-SDGs) indicators for the Sustainable Development Goals (SDGs)[3] and therefore countries around the world are expected to measure them in the near future. In this context, the research carried out for this paper has informed the development of methodologies for upgrading the indicators to be used for monitoring SDG 16, particularly targets 16.6.2 (proportion of the population satisfied with their last experience of public services) and 16.7.2 (proportion of the population who believe decision making is inclusive and responsive). It is also reflected in the Responsiveness chapter of the UN Praia City Handbook on Governance Statistics, an international collaborative initiative to guide the development of Governance Statistics that was endorsed by the United Nations Statistical Commission (UNSC) in April 2020.

Finally, another objective of this paper is identifying areas where, despite recent advances, room exists to improve the collection of these metrics. The remainder of this paper is structured as follows. Section 2 elaborates on the importance of these dimensions, and reviews the availability of existing statistics on external political efficacy and satisfaction with health care and education services. Based on existing evidence, Section 3 assesses the statistical accuracy of responsiveness metrics and sheds light on areas where room for improvement exists. Section 4 concludes and proposes steps that could be taken to move the statistical agenda ahead.

2 Importance of the dimension and available statistics

This section analyses in greater detail the theoretical importance of the responsiveness dimensions treated in this paper, and describes the different measurement efforts made for each of them. Each subsection, includes a summary table describing the statistics reviewed in this paper. Annexes I and II provide further details about the different survey instruments and their characteristics.

2.1 Political Efficacy: importance of the dimension

System responsiveness (or external political efficacy[4]) refers to people's feeling of having a say in what their government does. This is core to the legitimacy of public institutions, which owe their *raison d'etre* to the fact that they are acting on behalf of and for people. It is also of paramount importance to democratic systems, as it relates to the belief that political and social change are possible and that people can play a part in bringing about this change. This belief is also associated with the idea that it is worthwhile to perform civic duties and abide to political rules (OECD, 2017a). Persistently low or diminishing levels of system responsiveness will raise doubts in people's minds about whether governments are working for the interests of the majority and not just a few. In turn, resulting disenchantment could erode the foundations of democratic systems and nurture the emergence of populist responses (Geurkink et al., 2020).

In the academic literature, levels of system responsiveness are also related to citizen engagement, satisfaction with democracy and trust in public institutions (Denemark and Niemi, 2012; Ikeda et al, 2008; Uslaner and Brown, 2005). Finally, it is also acknowledged that system responsiveness can be built and destroyed by people's experiences when interacting with public institutions and also by institutions that are not perceived as being responsive to people's needs (e.g. policy-making processes and government decisions that do not respond to public preferences).

2.2 Political Efficacy: available statistics

There is a long tradition of collecting political efficacy measures through household surveys that started in 1952 with the inclusion of a battery of questions[5] in the American National Election Studies[6] (ANES) collected by the Survey Research Centre (SRC) and the Centre for Political Studies (CPS) at the University of Michigan. Originally, political efficacy was considered a unidimensional construct; however, researchers realized early on the existence of two distinct dimensions to the concept: internal efficacy (beliefs about one's own competence to understand and participate effectively in politics) and external efficacy (perceptions of governmental responsiveness) and tried to fit existing questions to those dimensions. Yet evidence on the reliability and validity of metrics included as part of this model was unsatisfactory (Morrell 2003).

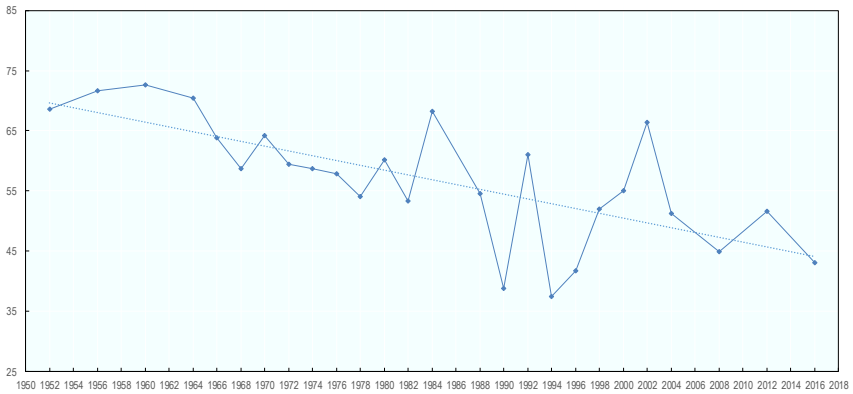
Still, from the onset of efforts to measure political efficacy, the so-called NOSAY question "People like me don't have a say about what the government does" has been included in surveys. An early wave of research associated the NOSAY question to internal political efficacy (Miller et al 1980; Craig and Maggiotto 1982; Acock and Clarke 1990). However, since 1987 new metrics[7] of internal political efficacy were introduced in ANES. Additional research showed that the NOSAY question is actually a measure of external political efficacy (Niemi et al 1991) a result confirmed empirically by subsequent waves of the survey and field experiments (Morrell 2003). Further research recognizes the NOSAY question as measure of external political efficacy (Borgonovi and Pokropep 2017, OECD 2017a).

The NOSAY question[8] has been included in the 2012 OECD Survey of Adult Skills (PIAAC) which was fielded in large and representative samples of the adult population in 32 countries and subnational entities. Since its seventh wave, fielded in 2014, the European Social Survey (ESS) is also collecting data on external political efficacy based on the following two questions: 1) "how much would you say the political system in [country] allows people like you to have a say in what the government does?"; and 2) "how much would you say that the political system in [country] allows people like you to have an influence in politics". Figures 2.1, 2.2 and 2.3 present data on external political efficacy from some of the surveys identified in this paper.

Figure 2.1 shows that the percentage of the American population who agrees with the statement "People like me don't have any say about what the government does" has declined by 25 percentage points, from 68.5% in 1952 to 43% in 2016. Originally the question was asked using a dichotomous scale ("agree" and "disagree") but since 1988 the response scale includes a third choice "neither agree nor disagree". The high levels of volatility of this measure suggest that this metric could be influenced by a range of factors such as the political cycle or the type of policies implemented by governments. However, the long term decline highlighted by Figure 2.1 is consistent with the erosion of institutional trust and social capital documented for the US over similar long term periods by OECD (2017a) and Putnam (2005)

Figure 2.1. Having a say in what the government does in the US

Percentage of the voting age population who feel they have a say in what the federal government does, 1952-2016



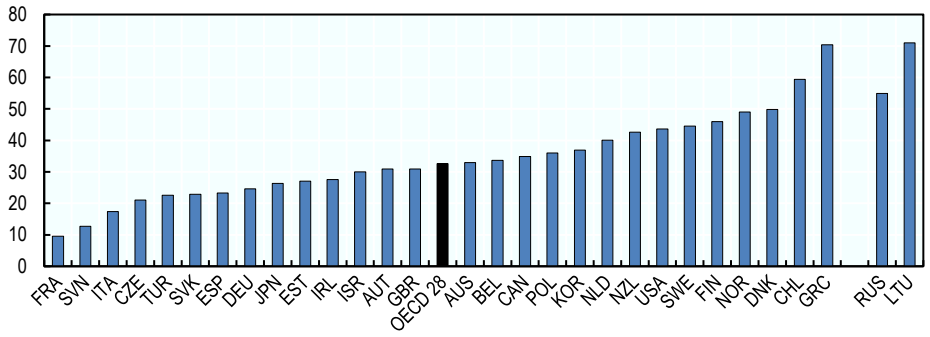
Note: The figure depicts the percentage of the population who agrees with the following statement: "people like me don't have any say about what the government does. Before 1988 the question had only two answer choices (agree, disagree and don't know). Since then the answer choice neither agree nor disagree was added.
Source: OECD calculations based on the American National Electoral Study combined dataset.

While the wording of the 2012 PIAAC questionnaire is identical to the ANES, PIAAC uses a 5 point Likert type scale ranging from 1 ("strongly agree") to 5 ("strongly disagree"). Comparative data for a meaningful set of countries is only collected every 10 years, with the next data collection planned for 2021-22.

According to the 2012 PIAAC data, one third of people in OECD countries believe having some influence on what government does, with the share ranging between 20% or less in Italy, Slovenia and France, to 60% or more in Chile, Greece and Lithuania (See Figure 2.2).

Figure 2.2. Having a say on what the government does

Percentage of the working-age population who feel they have a say in what the government does, around 2012



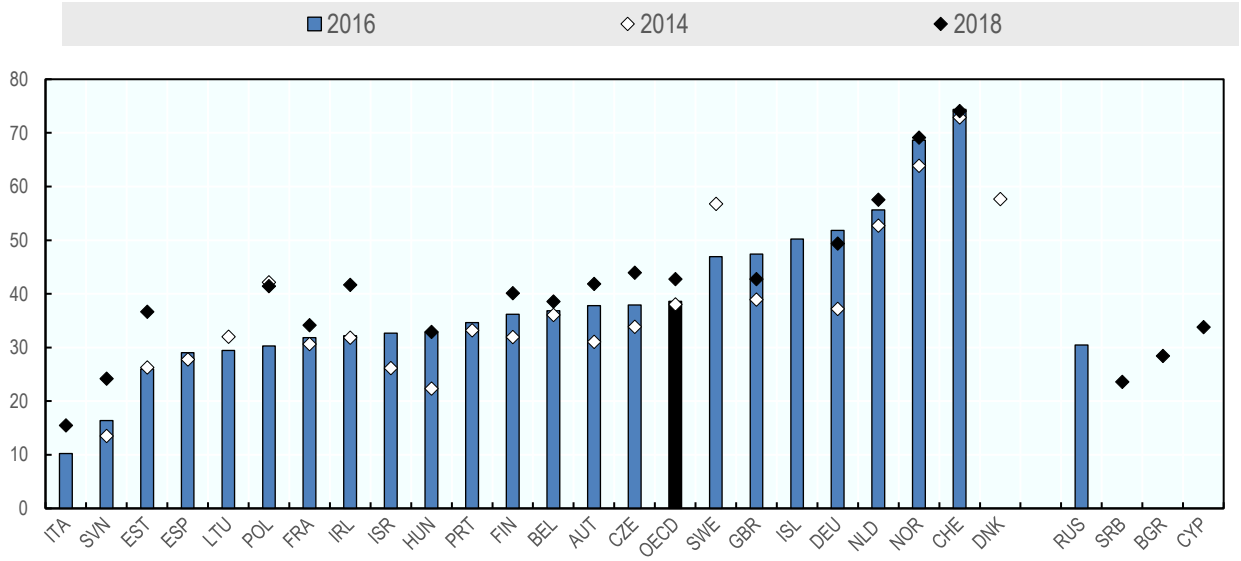
Note: The figure shows the combined share of people who "agree" or "strongly agree" with the statement "People like me don't have a say about what the government does". The higher the level the more people think they do have a say. Data refer to 2011-12 for Australia, Austria, Belgium, Canada, the Czech Republic, Denmark, Estonia, Finland, Germany, Ireland, Italy, Japan, Korea, the Netherlands, Poland, the Russian Federation, the Slovak Republic, Spain, Sweden, the United Kingdom and the United States; 2012 for France, and 2014-2015 for Chile, Greece, Israel, Lithuania, New Zealand, Slovenia and Turkey. Data for Belgium refers to Flanders; data for the United Kingdom refer to England and Northern Ireland; and data for the Russian Federation do not include Moscow municipal area. The OECD average is the simple average for the 28 countries with available data, and excludes Hungary, Iceland, Latvia, Luxembourg, Mexico, Portugal and Switzerland.
Source: OECD calculations based on data from OECD Survey of Adult Skills (PIAAC database), http://www.oecd.org/site/piaac/.

The ESS extended the measurement of political efficacy by adding from round seven (2014), when they were tested, questions on external political efficacy to the questions[9] on internal political efficacy included in rounds one (fielded in 2002) to five (fielded in 2010). One of the questions retained is: "How much would you say that the political system in [country] allows people to have a say in what the government does?" Respondents answered based on a five points Likert type scale[10] ranging from one (not at all) to five (a great deal).

Differently from ANES and PIAAC, the ESS question is worded in a positive way "have a say" instead of "don't have a say") asks about the political system rather than government, and the items are formulated as questions rather than as statements with which respondents can agree or disagree. According to (Saris and Torcal (2009) the use of questions instead of statements leads to higher quality data[11]. The ESS results show that, on average, about 42% of people in OECD countries with available information considered in 2018 having a say in what the government does (see Figure 2.3).

Figure 2.3. Having a say in what government does, 2014, 2016 AND 2018

Percentage of the population over 15 years who feel that the political system allows them to have a say in what the government does



Note: Data based on answers to the question "how much would you say the political system in [country] allows people like you to have a say in what the government does?" The data for 2016 and 2018 reflects the percentage who answered "some", "a lot" or "a great deal". Those for 2014 reflects the percentage of the population that answered five or more based on a scale from 0 ("not at all") to 10("completely agree").
Source: Author calculations based on waves 7, 8 and 9 of the European Social Survey (ESS)

Some OECD countries have also included questions on political efficacy in their official surveys. For example, the Mexican Ministry of Interior has conducted five rounds of a national survey on Political Culture and Citizenship (ENCUP), three of them in partnership with the Mexican National Statistical Office (INEGI). While the questions have varied, the topic of external political efficacy has been included in all waves of the survey. The most recent survey fielded in 2012 asked the following question: "Do you agree or disagree with this statement? People like me have influence on what the government does[12]", with answers based on a three points scale (See Annex I). Previous versions of the survey in 2003, 2005 and 2008 included the NOSAY question[13].

The two most recent waves of the Australian General Social Survey (AGSS) fielded in 2014 and 2010 included the following question: "How often do you feel you are able to have a say within the general community, on issues that are important to you?" In turn, their harmonised module on democratic governance of the Strategy for the Harmonisation of Statistics in Africa (SHaSA) asks the question "How often do you think the following listen to people like you?" with respect to the following institutions: "members of parliament/national assembly", "local elected officials/councillors" and "traditional leaders".

Political efficacy questions have also been included in other academic and research projects. For example, the Political Support in Canada Project led by an Inter University Consortium for Political and Social Research included several political efficacy questions (including the NOSAY one) in surveys conducted between 1983 and 1993. The Comparative Study of Electoral Systems (CSES) also includes two questions on external political efficacy (Karp and Banducci 2008). Finally, the International Social Survey Programme (ISSP) included in its quality of government module, fielded in five rounds between 1985 and 2016, a range of questions on political efficacy (including the NOSAY question). Table 2.1 presents the different sources on political efficacy questions identified in this paper, while Annex I provides more details about question wording and survey frequency.

Table 2.1. Household surveys including political efficacy questions

Source	Country coverage	Ongoing/ Suspended	Frequency political efficacy questions
American National Electoral Study (ANES)	1	Ongoing	1952-2018 linked to the electoral cycle (every 2 years)
European Social Survey (ESS)	38	Ongoing	Biennially since 2014
Survey of Adult Skills (PIIAC)	39	Ongoing	2012-2014 Every ten years
Political Support in Canada Project	1	Suspended	1983-1993 (three data points)
Survey on Political Culture and Citizenship-Mexico	1	Ongoing	2003, 2005, 2008,2012
Australian General Social Survey	1	Ongoing	Since 2002 every four years
Strategy for the Harmonization of Statistics in Africa	11	Ongoing	Linked to data collection cycles by NSOs
Comparative Survey of Electoral Systems	55	Ongoing	Staggered depending on elections Five modules from 1996-2021 1: 1996-2001 2: 2001-2006 3: 2006-2011 4:2011-2016 5:2016-2021
International Social Survey Programme	57	Ongoing	1985,1990,1996,2006,2016

Note: The table includes countries for which at least a data point exists

2.3 Satisfaction with education and healthcare services: importance of the dimension

The scope of government's action into different areas of service provision varies widely across countries, and the extent to which government is the main provider of services is a source of debate (Feldstein, 1996). Still, satisfaction with government services is an outcome of government activity that captures elements that are essential to people's lives (OECD, 2017a; OECD, 2017b). In many instances the performance of public institutions has been associated with the quality of public services, traditionally measured through user satisfaction (Van Ryzin, 2004; Van Ryzin, 2006). This is fundamental as people expect high returns for their tax payments, mainly in the form of service provision (Van de Walle, 2018). This paper addresses particularly healthcare and education services and therefore is not representative of all government services. Nonetheless, education and healthcare are not only the services more widely covered by household surveys but alongside social protection represent the bulk of government service provision activities in terms of staff and spending (OECD, 2019a).

In addition, the opportunities available to people for improving their quality of life depend upon access to collective as well as individual resources, in areas that are key for building human capital such as education and healthcare among others. Along these lines, policy measures to address persistent or deteriorating economic, social and employment problems emphasize the role of social and other public services in building capabilities and levelling up the field in access to opportunities (OECD 2019c; OECD 2019d)

A positive relation between perceived performance and user satisfaction has been established in the literature. The mechanism linking performance and satisfaction is known as "expectancy disconfirmation theory"[14]. According to this theory, citizens evaluate government services through a combination of experience and expectations (Young Mok et al., 2017; Oliver 2010). Expectations are formed as a summative judgement of past experiences[15], including experiences with private service provisions and assessments by others, and shape users' satisfaction with a given service.

People may have direct experiences of certain services while using sporadically (or never) others. Lack of direct experience do not necessarily refrain people (i.e. survey respondents) from answering satisfaction questions about both types of services (i.e. experienced and not experienced).Additionally, research has documented the existence of the so called *halo* effect, the notion that attitudes towards the public sector could influence people's answers to questions about satisfaction with public services (Van de Walle, 2017). This underscores the importance of asking detailed questions not only about the use of services but also on specific service attributes such as the easiness of access and quality that are considered as determinants of satisfaction levels (OECD, 2019a).

In turn, satisfaction with services shapes political attitudes and behaviours towards policies (Young Mok et al, 2017). People who are dissatisfied with the public services they consume are likely to demonstrate their dissatisfaction by raising their voice, filing a complaint, engaging in protest or opting out of public provision if private providers are available (Van de Walle, 2018). Also improving the quality of public services can lead to more satisfied users, which, in turn, can increase trust in government institutions (another key outcome of government activities), a mechanism referred in the literature as the "micro-performance hypothesis" (Van de Walle and Bouckaert, 2003; Yang and Holzer, 2006).

All in all, three hypothesis stem from the academic literature as having an effect in satisfaction with services. These hypothesis are: 1) the existence of a *halo* effect, i.e. The influence that attitudes about the government or public sector could have in shaping people's assessment of the public services they use. 2) The role of people's *experiences* as a driving force shaping their satisfaction (expectancy disconformity theory) and 3) the extent to which satisfaction is influenced by various *features of the service*, such affordability, proximity, quality etc.

2.4 Satisfaction with education and healthcare services: available statistics

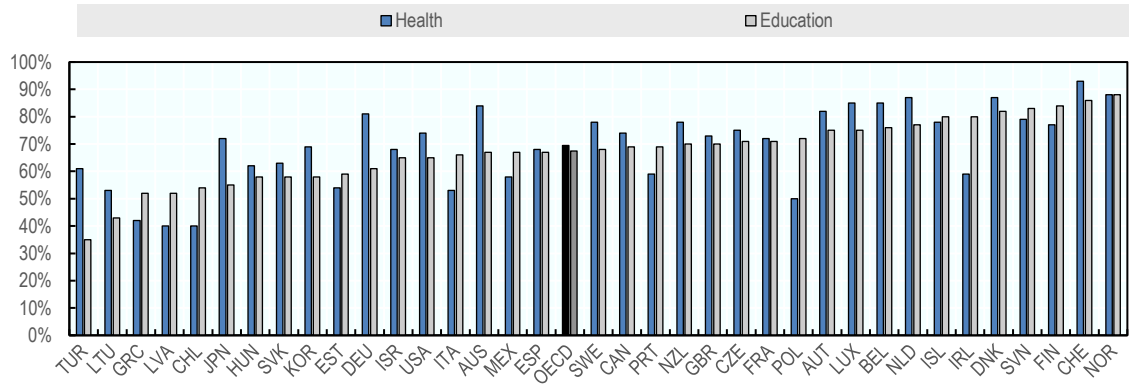
Public managers have long experience with surveys specifically focused on measuring users' satisfaction with services, conducted mainly at the local level and with varying degrees of representativeness (Bouckaert and Van de Walle, 2003). In addition, a non-comprehensive review found that 11 NSOs[16] including Mexico, Peru, Brazil, Germany, Norway and South Africa – have included questions about satisfactions with services in their regular data collections, although with different scope and frequency (UNDP 2017) and with differences in terms of the services covered[17] and the type of questions asked. Health-care and education services are those most commonly covered, and are the focus of this section[18].

Non-official household surveys have also included questions on public services, with some surveys asking about satisfaction with services in the city or area of the respondent (e.g. Gallup World Poll) while others ask about satisfaction with the state of the health and education systems in the respondent's country (e.g. ESS, EQLS). A common feature of these surveys is that they do not distinguish between services provided publically or privately rendering impossible a sectorial analysis.

Given its coverage, the Gallup World Poll (GWP) is often used source for measuring satisfaction with services (OECD 2017b). Figure 2.5 shows the percentage of the population in OECD member and accession countries who are satisfied with the availability of health care and or schools at the local level, with answers reported based on a binary choice (i.e. satisfied/dissatisfied). According to the latest available data, 67% of the population in OECD countries is satisfied with the education system while 69% is satisfied with the availability of healthcare.

Figure 2.4. Satisfaction with health and education in OECD member countries, 2019

Percentage of the population who are satisfied



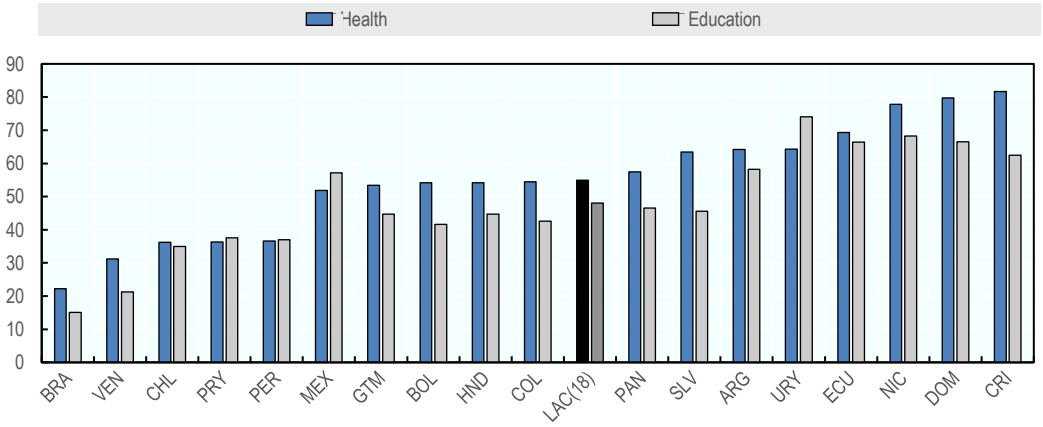
Note: Percentage of the population who answered "yes" to the question: "in the city or area where you live how satisfied you are with" "the availability of healthcare" and "the education system or schools"
Source: OECD calculations based on the Gallup World Poll.

A slightly different approach is used by the Latinbarometer, which restricts the questions on health and education to the services respondents have access to (See Annex II for details of the questions included in cross-country comparative surveys discussed in this paper). Other surveys such as the ISSP ask respondents about their desired level of spending in key services, as well as who should be providing these services (e.g. public, private or mixed provision) but does not include questions about satisfaction.

Annual[19] data from Latinobarometer on satisfaction with several services[20] are available since 1995 for eighteen LAC countries. Since 2016 this survey asks questions on both satisfaction and on having specific access[21] to those services. Latinobarometer also includes questions about specific experience with public hospitals and schools. According to the latest available data, 55% of the population in LAC countries are satisfied with the health system they have access to, and slightly less than half report the same with the education system (see Figure 2.6).

Figure 2.5. Satisfaction with health and education in Latin America, 2016

Percentage of the population who are satisfied with the education and health they have access to



Note: Percentage of the population who answered very satisfied or rather satisfied to the question: Would you say that you are very satisfied, rather satisfied, not very satisfied or not at all satisfied with...? Your access to health, your access to education.
Source: OECD calculations based on Latinobarometer

An alternative approach consists of breaking the satisfaction measure into different components (i.e. operational qualities) that may influence satisfaction. For example, the OECD's Serving Citizens framework[22] evaluates three key public services (i.e. health care, education and justice) based on nine sub-dimensions including, among others, geographical and financial access, timeliness (e.g. time for initiating a surgery after admission to the hospital), courtesy, treatment (e.g. patient contact with doctor and involvement in decision of treatment) and quality of the services provided. A key advantages of this approach is that it acknowledges the variety of factors that could shape users' satisfaction. However, a big challenge lies in determining how each of these factors influences overall satisfaction. Furthermore, sub-dimensions for different services cannot be measured in the same way. For example while quality in the health sector could be measured by mortality rates after undertaking a specific type of surgery, quality of educational services may be measured by scores in standardized tests (e.g. PISA or PIRLS) that in turn could be influenced by factors unrelated to the education system such as different contexts at home.

A combined approach has been used in the field of health by the OECD Patient-Reported Indicators Survey (PaRIS), launched in 2017. This instrument includes Patient-Reported Experience Measures (PREMS), measuring how patients experience care in practical terms (e.g. waiting time, doctor-patient communication etc,) and Patient Reported Outcome Measures (PROMS), which measure how patients assess the results of the care they receive (e.g. satisfaction with the outcomes of a specific procedure such as breast cancer surgery). This approach goes a step forward in assessing operational qualities at the level of very specific interventions. Table 2.7 presents the different sources providing information on satisfaction with services, while Annex II provides details about question wording and survey frequency.

Table 2.2. Surveys including satisfaction with services questions

Source	Country coverage	Ongoing/Suspended	Frequency, satisfaction with services questions
Gallup World Poll	134	Ongoing	Annual since 2006
The Quality of Government Index Survey (EQI)	27	Ongoing	2010,2013 and 2017
Latinbarometer	18	Ongoing	Annual since 1995*
European Social Survey	38	Ongoing	Biennially since 2014
European Quality of Life Survey	34	Ongoing	2003,2007,2012,2016

Note: The table includes countries for which at least a data point exists

* Data not collected in 1999.

3 Accuracy of responsiveness measures

This section assesses the statistical accuracy of metrics of external political efficacy and satisfaction with health care and education services as collected in household surveys. By doing so it closes a gap in the academic literature and contributes to the development and refinement of methodologies for measuring these aspects in the future, for example in the measurement of SDG 16 relevant targets.

The accuracy of a metric is the degree to which it captures the concept that it is intended measure. Accuracy is typically thought of as having two dimensions: reliability and validity. The reliability of a metric is the degree to which repeated measurements of the same concept thing produce the same results. In this sense, a reliable measure is one that involves minimal “noise” in the measurement process. Validity, on the other hand, is concerned with whether the measure in question is biased. A valid measure is one that reflects the underlying concept being measured with no systematic bias.

A measure can be reliable while having low validity if it is not noisy but bears little direct relationship to the concept being measured. At the limit, a highly reliable but invalid measure will be “precisely wrong”. At the other extreme, a measure can be valid without being very reliable. Such a measure will be accurate on average, but individual measurements will have a high level of noise. An accurate measure needs adequate levels of both reliability and validity, although the desirability of different trade-offs between the two will depend on the precise use to which the data is to be put.

Validity is usually analysed in terms of face validity (whether the measure makes sense intuitively), convergent validity (whether the measure correlates well with other proxy measures of the same concept) and construct validity (whether the measure behaves as theory and common sense dictate). A measure can be considered valid if it performs well in terms of all three aspects of validity outline mentioned above[23]. Ideally, this would be assessed through a large and well-developed scientific literature that covers the main type of validity, and which is sufficiently mature for a consensus to have emerged. The empirical assessment presented below contributes to the evidence needed to inform the development and refinement of the metrics on these aspects, particularly SDG 16 targets.

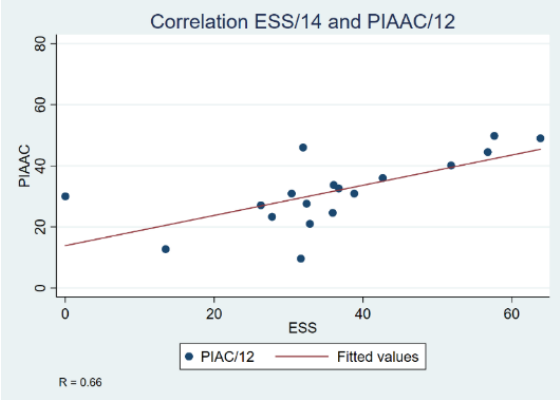
3.1 Reliability of political efficacy measures

A reliable measure produces consistent results when measured at similar moments by different instruments or when the same respondent is asked the same question more than once (test-retest reliability)[24]. In survey research, the standard measure of reliability is a test-retest reliability, where the same measurement item is administered to the same person after some delay. This may be later, after a fixed period, on in the same survey at different places. Few studies have assessed the test-retest reliability of political efficacy questions. McPherson Welch and Clark (1977), who used path analysis[25] to estimate

item reliability of the political efficacy construct, concluded that the NOSAY question is one of the items with highest reliability among political efficacy questions.

In the absence of further test-retest evidence, another method for assessing the reliability of external political efficacy measures at the country level is possible. When different measures of the concept are available for a wide number of countries using different survey instruments, it is possible to look at cross-country correlations between different surveys fielded in the same year or in 2 close points in time. Based on this method, Figure 3.1 shows the correlation between round 7 of the ESS, fielded in 2014, and the PIAAC collection undertaken around 2012. The correlation coefficient of 0.66, which is relatively high considering the lag between the two surveys.

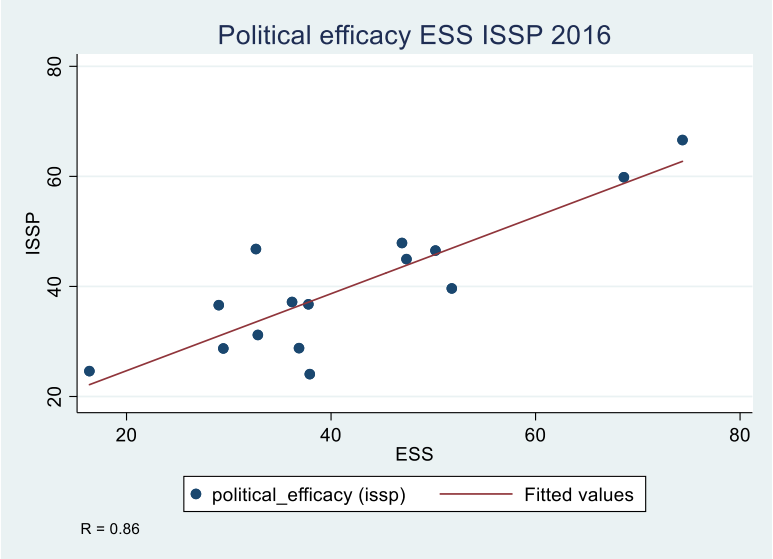
Figure 3.1. Cross-country correlation of political efficacy as measured in PIAAC and ESS, around 2014



Note: PIAAC data refer to the share of people who "disagree" or "strongly disagree" with the statement "People like me don't have any say about what the government does"; the higher the level, the more people think they do have a say. Data refer to 2011-12 for Austria, Belgium, the Czech Republic, Denmark, Estonia, Finland, Germany, Ireland, Italy, the Netherlands, Poland, the Slovak Republic, Spain, Sweden and The United Kingdom. 2012 for France; and 2014-15 for Slovenia. Data for Belgium refers to Flanders; data for the United Kingdom refer to England and Northern Ireland. ESS data are from round seven (2014) and refer to the percentage of the population who answered the question ": how much would you say the political system in [country] allows people like you to have a say in what the government does?" with a score of five or more on a (0-10 scale).
Source: OECD calculations based on round 7 of the ESS and PIAAC.

Similarly, Figure 3.2 shows the cross-country correlation between the NOSAY question in ESS and that in the International Social Survey Programme (ISSP). The latest collection in IPPS took place in 2016, the same year of round 8 of the ESS. The correlation coefficient (0.86) in this case is higher than the ones shown in Figure 3.1).

Figure 3.2. Cross-country correlation between measures of political efficacy in ISSP and ESS, 2016



Note: ISSP data refers to the percentage of the population who "strongly disagree", "disagree" or "neither agree nor disagree" with the statement "people like me don't have any say about what the government does". The ESS refers the percentage who answered "some", "a lot" or "a great deal" to the question: "how much would you say the political system in [country] allows people like you to have a say in what the government does?"
Source**:** OECD calculations based on round 8 of the ESS and wave 5 of the ISSP

It is also possible to examine the reliability of the measure over time. Table 3.3 displays the intra-wave correlation in the European Social Survey (ESS) of political efficacy measures. As discussed previously, political efficacy questions were tested in 2014 before being included in the core questionnaire in 2016 and 2018. Despite changes in the response scale, the correlation of 0.97 signals strong reliability of the measure.

Table 3.1. Intra-wave correlation between measures of political efficacy in waves 7, 8 and 9 of the European Social Survey

	Wave 7 (2014)	Wave 8 (2016)	Wave 9 (2018)
Wave 7 (2014)	1		
Wave 8 (2016)	0.91	1	
Wave 9 (2018)	0.97	0.97	1

Source: OECD calculations based on waves 7, 8 and 9 of the European Social Survey
Note: Data are from rounds 7 (2014) and 8 (2016) of the European Social Survey, and refer to the question: "how much would you say the political system in [country] allows people like you to have a say in what the government does?". Data for 2016 refers to the percentage who answered "some", "a lot" or a "great deal". Those for 2014 reflects the percentage of the population who answered five or more (on a 0 to 10 scale). Data for wave 9 were extracted in February 15 2020 and include 20 countries.
Source: OECD calculations based on rounds 7, 8 and 9 of the ESS.

It is also possible to test the intra-wave correlation for the ISSP which has collected the NOSAY question every ten years since 1996. Table 3.4 displays the correlation coefficient across waves. These correlations are high but tend to decrease as rounds are far apart, reflecting changes the assessment of the concept by respondents. All in all, and despite the limited number of sources available, the evidence in this section suggests a high reliability of the NOSAY/SAY question. This evidence is consistent with arguments made

by early theorist of political efficacy who considered this concept as deep-seated orientation, rather than a volatile attitude (Campbell et al 1954, Lane 1959).

Table 3.2. Intrawave correlation between waves 3, 4 and 5 of the International Social Survey Programme (ISSP)

	Wave 3 (1996)	Wave 4 (2006)	Wave 5 (2016)
Wave 3 (1996)	1		
Wave 4 (2006)	0.67	1	
Wave 5 (2016)	0.66	0.83	1

Note: The ISSP refers to the percentage of the population who "strongly disagree", "disagree" or "neither agree nor disagree" to the statement "people like me don't have any say about what the government does".
Source: OECD calculations based on waves 3, 4 and 5 of the International Social Survey Programme.

Taken together, Tables 3.3 and 3.4 support the view that the NOSAY question is reliable at the country-average level. Where the question is directly comparable in surveys of relatively high quality, the cross-country correlations are in excess of 0.80. Even when the lag between surveys is close to 10 years, the correlation is still above 0.6. This evidence indicates that, on the basis of available information, the NO-SAY question is amenable for inclusion in surveys interested in measuring system responsiveness.

3.2 Validity of political efficacy measures

3.2.1 Face validity

Face validity addresses the degree to which a measure is intuitively plausible. This is a subjective judgment, and not one for which a definitive answer is possible. Some objections to the wording of the NOSAY/SAY question have been raised. For example, it has been argued that the negative wording of the question could prompt respondents to accentuate the negative, and that "acquiescence bias[26]" could lead to reported high levels of political inefficacy. (Clarke et al 2010) have used a split sample experiment in the Political Support in Canada (PSC) study to analyse the effects of negatively worded questions versus positive worded ones, and found no evidence that negatively framed statements led to acquiescence bias, and that these performed better in a highly contested election Still, data on political efficacy based on positively worded question are easier to communicate, which led the ESS to include a positive formulation of the question, founding no evidence that negative wording outperforms the positive one (Saris and Revilla 2012 and Saris and Torcal 2009).

A more quantitative approach is to look at how difficult respondents find it to answer questions on external political efficacy. A simple way of evaluating this is to look at the item-specific non-response rates for measures of political efficacy compared to other widely used survey items. Figure 3.5 shows the item-specific non-response rates in rounds seven, eight and nine of the European Social Survey. These rounds included questions on satisfaction with democracy and institutional trust as well as on topics unrelated to political efficacy commonly asked in surveys such as income, education, religion, employment, gender and age. Household income, is by a large margin the most difficult or sensitive topic for survey respondents given the multiplicity of sources and recall problems or confusion (Moore et al 2000). About 20% of respondents either refuse to answer or respond with "don't know" to questions about income. At the other end of the spectrum, gender and age have extremely small "don't know" response rates.

The ESS question on political efficacy performs better than those on income, slightly better than that on satisfaction with democracy, and similarly to that on trust in parliament. Still, the share of "don't know" (2.2%) is higher than for questions about trust in law and order institutions (i.e. the police), life satisfaction

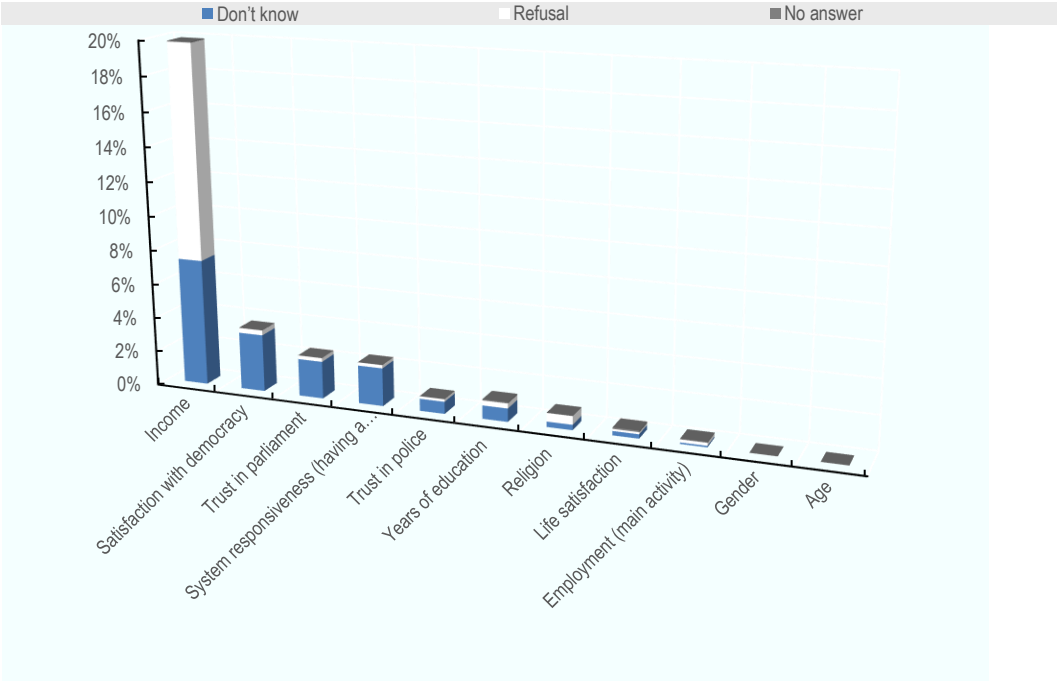
and objective items such as years of education and age. This suggest that, although the question is easy to understand, some people lack adequate information to provide a meaningful answer.

In early 2018, the World Values Survey (WVS) agreed to test the NO-SAY question in 15 Non-European primarily developing countries. The average non-response rate for these countries was 5.6%, which is higher than for the European countries covered by ESS (UNDP 2019a). The different non-response rates between surveys suggests that asking about the government is more sensitive in some countries than in others, which might make some respondents more reluctant to answer.

The overall picture on face validity emerging from an analysis of non-response rates is, hence mixed. While the NOSAY question perform better than income – an item which is commonly collected in almost all household surveys - the item-specific non-response rates for the political efficacy question are higher than for questions on gender, employment and education. Even religion, which might be considered a relatively sensitive question, had a non-response rate that was significantly smaller than the NOSAY question.

An analysis of non-response rates does not, in itself, provide conclusive evidence on face validity. Relatively high non-response rates for political efficacy could simply reflect that although the question is reasonable, many people lack adequate answer to provide a meaningful answer. The higher non-response rate for concepts that are more abstract such as trust in parliament and satisfaction with democracy, provide support for this view.

Figure 3.3. Item-specific non-response rates in the European Social Survey, 2014, 2016 and 2018

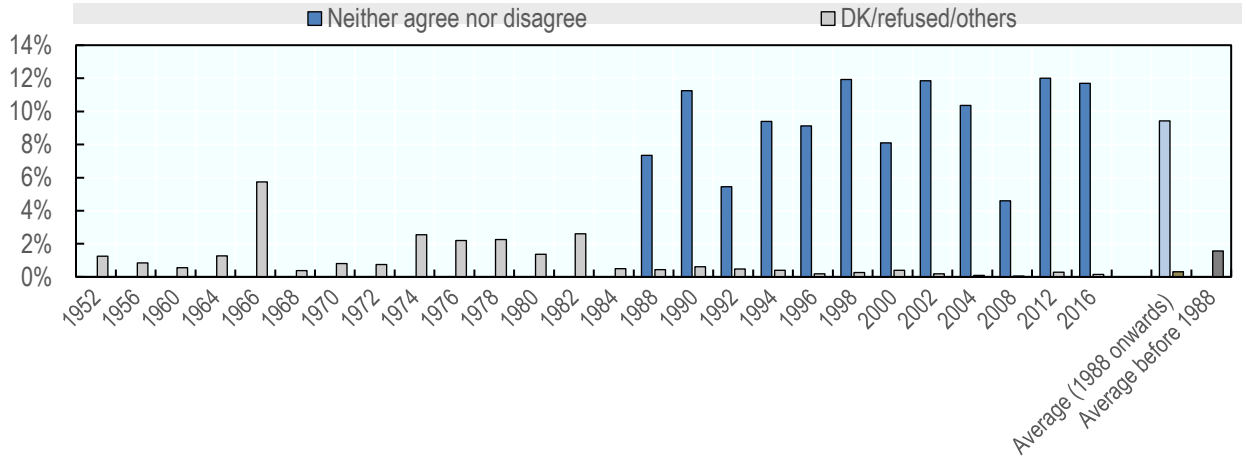


Source: OECD calculations based in rounds seven and eight of the European Social Survey

The American ENAS provides a natural experiment for assessing the importance of non-response rates in surveys including political efficacy questions. Until 1988 the response categories for the ENAS NOSAY question were "agree", "disagree", "don't know" and "refused". From 1988 onwards, the response option "neither agree nor disagree" was included. Figure 3.6, plotting item-specific non-response rates before and after 1988, show that these rate declined from 1.8% to 0.3%. In turn, the average share of respondents

who selected the "neither agree nor disagree" reached 9.4% after 1988. This suggests that surveys including political efficacy questions allow for an intermediate choice as part of their response scale. By doing so, respondents won't be pushed into one of the positive or negative response categories, and fewer will refrain from answering the question.

Figure 3.4. Item specific non-response rates for questions on political efficacy before and after the introduction of a neutral category of response in ENAS



Source: OECD calculations based in ENAS

3.2.2 Construct validity

Construct validity assesses whether a measure behaves as theory and common sense dictate. In a mature democratic system, people are expected to be confident of their ability to exert political influence (i.e. internal efficacy), and this expectation should be rooted in the realities of system performance (i.e. system responsiveness/external political efficacy). While the relationship is not necessarily causal, higher levels of political efficacy should be associated with higher civic and political participation as well as higher institutional trust[27].

Both political efficacy and trust have declined over time although the reasons for such a fall are different in the two cases (Norris 2016). Based on data from the 1972 CPS American Election Study, Craig (1979) provided evidence that external efficacy and trust were related but intrinsically different concepts. The academic literature argues that, while external efficacy measures a respondent's belief that government is responding to citizens' expectations regardless of the quality of the outputs, trust is associated with a normative belief about the quality of the outputs (Pollock, 1983; Hetherington, 1998; Chamberlain, 2012).

The OECD Guidelines on Measuring Trust (OECD 2017c) suggest that several dimensions of trust that can be captured by different questions. Using data from rounds seven, eight and nine of the ESS, and following the proposal put forward by the OECD Guidelines[28], Table 3.7 evaluates the pairwise correlation between measures of system responsiveness (as captured by the NOSAY question), trust in parliament (political institutions) and trust in the police (law and order institutions). While both correlations are statistically significant, the correlation of political efficacy is higher (0.80) when assessed with respect to trust in political institutions than with trust in the police.

Changes in political efficacy also affect political participation (Karp and Banducci, 2008). (Ferrín 2016) has shown that the question "as a whole, how satisfied are you with the way democracy works in your country" provides a reliable measure of respondent's valuation of how well the liberal elements[29] of democracy

work. Table 3.7 shows a high and statistically significant correlation between measures of political efficacy and of satisfaction with democracy. Conversely, the pairwise cross-country correlation between system responsiveness and of self-reported voter turnout is small and non-significant. However, this latter result should be taken cautiously due to the large gap between voting patterns as reported in surveys and actual turnout (OECD 2017a; IDEA 2016; OECD 2015).

Research investigating the relationship between direct democracy mechanisms and political efficacy has found mixed results; while Bowler and Donovan (2002) and Smith and Tolbert (2004) have found that direct participation mechanisms enhance political efficacy Dyck and Lasher (2008) concluded that this relations is either very weak or non-existent. The absence of questions on possibilities for direct participation[30] in rounds seven, eight and nine of the ESS does not make it possible to test this relation.

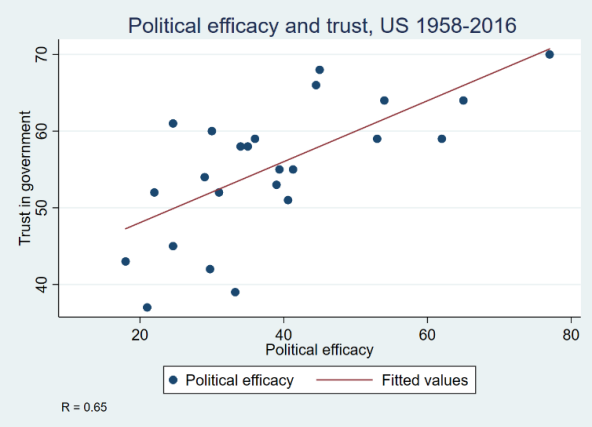
Table 3.3. Pairwise correlations between measures of political efficacy and other governance variables

	Political efficacy	Satisfaction with democracy	Trust in parliament	Trust in the police	Self-reported voter turnout
Political efficacy (SAY)	1				
Satisfaction with democracy	0.82**	1			
Trust parliament	0.80**	0.93**	1		
Trust police	0.52**	0.69**	0.76**	1	
Voted	0.28	0.43**	0.50**	0.40**	1

Note: ** indicates that the correlation is statistically significant at 99%.
Source: European Social Survey, rounds seven, eight and nine.

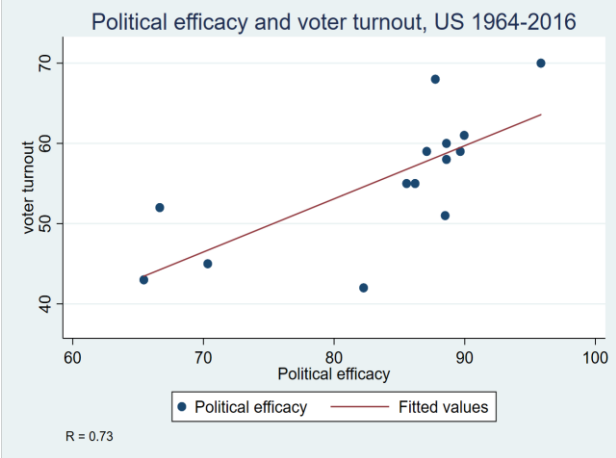
Analysis on construct validity based on a single source could be criticised as being affected by shared method variance (Jakobsen and Jensen 2015). While no other cross country comparative survey has been identified as including similar questions, US data allow investigating relations over time. Figure 3.8 shows the correlation between the NOSAY questions asked in the ANES and the trust in the national government question from the PEW Research Centre between 1958 and 2016. The correlation is positive and significant. Figure 3.9 also presents the correlation between survey measures of political efficacy and administrative data on voter turnout in presidential elections as reported in the Institute for Democracy and Electoral Assistance (IDEA) dataset; differently from what observed in the ESS this correlation is high (0.73) and statistically significant.

Figure 3.5. Political efficacy and trust in government in the United States, 1958-2016



Note: Trust data refers the percentage of the population who "trusts the government in Washington to do what is right". Prior to 1985 Trust data were not collected on an annual basis. Political efficacy refers to the percentage of the population disagreeing with the statement "people like me don't have any say about what the government does". Before 1988 the question had only three response categories (agree, disagree and don't know); since then the response category "neither agree nor disagree" was added.
Source: Trust data is from the PEW research Centre, Political efficacy data is from the ANES

Figure 3.6. Political efficacy and voter turnout in the United States, 1964-2016



Note: Voter turnout rate is the ratio between the number of votes recorded by electoral authorities and the voting age population. Turnout data refers to presidential elections. Political efficacy refers to the percentage of the population disagreeing with the statement "people like me don't have any say about what the government does", Before 1988 the question had only three response categories (agree, disagree and don't know); since then the response category "neither agree nor disagree was added.
Source: Turnout data is from IDEA, political efficacy is from ANES

Generally speaking", measures of system responsiveness are well supported in terms of construct validity. While only few papers have analysed external political efficacy from the perspective of construct validity, a large literature has looked at its determinants and correlates. This reflects a high degree of interest in measures of political efficacy within political science, public management, sociology, psychology and economics. Indeed, the fact that political efficacy measures perform so well in terms of construct validity helps to explain why there has not been a stronger emphasis on more formally testing other forms of validity of such measures.

3.2.3 Convergent validity

A measure is said to have convergent validity if it correlates well with other measures of the same construct. Historically, external political efficacy has been measured with a battery of questions implying the existence of redundancies. Moreover, most of the research has been directed towards identifying the questions with the highest discriminatory power. Several studies have analysed the discriminatory power of diverse batteries of political efficacy questions, concluding that the NOSAY and influence in politics questions have the highest discriminatory power (Acock and Clarke, 1990; Craig et al., 1990; Niemi et al., 1991; Morrell, 2003; Saris and Revilla, 2012)

Table 3.10 presents the pairwise correlations of all external efficacy/system responsiveness questions included in the European Social Survey. All correlations are relatively high and statistically significant, especially those that are closer to the spirit of the NOSAY question, e.g. “have influence in politics” and “politicians care about what people think”. Confirmatory factor analysis[31] among these data confirm the goodness of fit of the single factor model; standardized factor loading are highest for the questions on the extent to which the political system allow people to have influence on politics and whether people considers having a say in what the government does (See Annex III). All in all, there is strong evidence of convergent validity for the NOSAY question.

Table 3.4. Cross-country pairwise correlation between measures of political efficacy and proxy concepts

	SAY	APG	HIP	PPT	EPP
Have a say in what the government does? (SAY)	1				
Have influence in politics? (HIP)	0.92*	0.90*	1		
Politicians care what people think? (PPT)	0.83*	0.83*	0.95*	1	
Easy to take part in politics (EPP)	0.79*	0.94*	0.92*	0.87*	1

Source: OECD calculations based on round 7 of the ESS

Table 3.11 summarises the evidence presented in this section by providing a snapshot of the quality of political efficacy questions based on the different criteria for statistical quality considered in this paper. Our assessment, based on available evidence, is that the accuracy (i.e. reliability and validity) of measures of system responsiveness is generally robust. Hence a strong case exists to promote the collection of these statistics. Still, there is room for improvement by further investigating and enhancing the performance of these metrics. For example it would be important to test directly whether the selection of different wording by different surveys (e.g. *political system* as finally decided in ESS and *government* in PIAAC and ISSP) has any impact on the results or if these terms are fully interchangeable (e.g. they capture the same concept). It will also be desirable to agree on a common response scale or a system of equivalences between the different scales currently being implemented in different surveys.

Testing how political efficacy questions are influenced by the closeness of elections could provide additional valuable information. While the ANES links its data collection to the electoral cycle other surveys such as ESS and ISSP follow their own internal data collection cycles, a plausible hypothesis is that people will report higher levels of political efficacy the closer the measurement is to the elections. Understanding the existence and size of this *closeness to elections* effect could shed light on what are the structural levels of the political efficacy metric and what are possible cyclical effects. In turn, there is also room for assessing the extent to which new digital technologies and additional, more frequent channels of interaction, participation and engagement between people and the public administration are affecting external political efficacy levels.

Table 3.5. Summary evaluation of the accuracy of survey measures of system responsiveness

	Face validity	Construct validity	Convergent validity	Reliability
NOSAY question	√√	√√	√√	√√

Note: √√ mean strong validity or reliability while √ means weak validity or reliability

3.3 Reliability of satisfaction with education and healthcare measures

In order to test the reliability (i.e. the degree to which a measure produces consistent results when measured at different times) of measures of satisfaction with health care and education, we constructed a cross country comparative panel dataset spanning from 2006-2017. The dataset includes measures of satisfaction with services as well as other relevant covariates from five cross country comparative non-official household surveys: the Gallup World Poll, The Quality of Government Index Survey, the Latinbarometer, the European Social Survey and the European Quality of Life Survey). Annex II provides details on the satisfaction questions included.

Tables 3.12 and 3.13 present pairwise cross-country correlations for metrics of satisfaction with services from these different sources. With respect to health care, all correlation in Table 3.12, are at or above 0.80 signalling strong reliability of measures of satisfaction with health; the only exception is the correlation between the Gallup World Poll and Latinbarometer, which is lower (0.73) but still statistically significant.

Table 3.6. Cross-country pairwise correlation between measures of satisfaction with health care services from different household surveys, 2006-2017

	Gallup	QoG	Latinobarometer	ESS	EQLS
Gallup	**1**				
QoG	0.87*	**1**			
Latinobarometer	0.73*		**1**		
ESS	0.80*	0.90*		**1**	
EQLS	0.83*			0.84*	**1**

Note: *Correlation statistically significant at 99%.
Source: OECD calculations based on the satisfaction with services dataset.

In the case of satisfaction with education services, the correlations shown in Table 3.13 are lower than those reported for health care and, for the one between QoG and ESS, non-statistically significant. A plausible explanation for these lower correlations is that while all segments of the population are using health services, most education services are restricted to a specific age group, implying that segments of the populations without direct exposure are being asked to pass judgment on a service based on distant experiences or referred knowledge. Even if the case of direct exposure (i.e. parents with kids), the experience tends to be indirect, as it is based on reporting by children who attend schools. These lower coefficients overall signal weaker reliability of measures of satisfaction with education.

Table 3.7. Cross-country pairwise correlation between measures of satisfaction with education services from different household surveys, 2006-2017

	Gallup	QoG	Latinobarometer	ESS	EQLS
Gallup	1				
QoG	0.65*	**1**			
Latinobarometer	0.75*		**1**		
ESS	0.71*	0.53		**1**	
EQLS	0.73*			0.74*	**1**

Note: *Correlation statistically significant at 99%.
Source: OECD calculations based on the satisfaction with services dataset.

The reliability of the measures over time can also be analysed. Tables 3.14 to 3.19 show the intra-waves correlations for measures of satisfaction with healthcare and education between the different non-official surveys included in the dataset. Correlations between the nearest waves tend to be higher while they decrease as waves are far apart, a result which indicates that the metric captures slow changes in the assessment of the services. Intra wave correlations are also consistently smaller in the case of education than for health care.

Table 3.8. Intrawave correlation of measures of satisfaction with health services between waves 1 to 8 of the European Social Survey

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Wave 1 (2002)	1							
Wave 2 (2004)	0.95	1						
Wave 3 (2006)	0.96	0.87	1					
Wave 4 (2008)	0.90	0.82	0.96	1				
Wave 5 (2010)	0.86	0.71	0.94	0.95	1			
Wave 6 (2012)	0.89	0.80	0.91	0.92	0.93	1		
Wave 7 (2014)	0.84	0.88	0.84	0.87	0.78	0.93	1	
Wave 8 (2016)	0.85	0.75	0.88	0.89	0.87	0.91	0.96	1

Note: Percentage of the population who answered five or more to the following question: please say what you think overall about the state of health services in [country] nowadays?
Source: OECD calculations based in the Satisfaction with Services dataset

Table 3.9. Intrawave correlation of measures of satisfaction with education services between waves 1 to 8 of the European Social Survey

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
Wave 1 (2002)	1							
Wave 2 (2004)	0.88	1						
Wave 3 (2006)	0.90	0.97	1					
Wave 4 (2008)	0.88	0.84	0.96	1				
Wave 5 (2010)	0.80	0.71	0.92	0.95	1			
Wave 6 (2012)	0.83	0.84	0.88	0.87	0.88	1		
Wave 7 (2014)	0.85	0.85	0.86	0.83	0.84	0.95	1	
Wave 8 (2016)	0.77	0.73	0.83	0.80	0.89	0.84	0.95	1

Note: Percentage of the population who answered five or more to the following question: please say what you think overall about the state of education services in [country] nowadays?
Source: OECD calculations based in the Satisfaction with Services dataset

Table 3.10. Intrawave correlation of measures of satisfaction with health services between different waves of Latinbarometer

	2007	2009	2010	2011	2015	2016
2007 Round	1					
2009 Round	0.69	1				
2010 Round	0.62	0.89	1			
2011 Round	0.51	0.77	0.94	1		
2015 Round	0.53	0.72	0.82	0.80	1	
2016 Round	0.61	0.73	0.74	0.70	0.8093	1

Note: Percentage of the population who answered satisfied or rather satisfied to the following question: Would you say that you are very satisfied, rather satisfied, not very satisfied or not all satisfied with: Health care to which you have access? Years between 2006 and 2017 that are not included indicate that the survey was not conducted or the question were not asked.
Source: OECD calculations based in the Satisfaction with Services dataset

Table 3.11. Intrawave correlation of measures of satisfaction with education services between different waves of Latinbarometer

	2007	2009	2011	2015	2016
2007 Round	1				
2009 Round	0.65	1			
2011 Round	0.55	0.87	1		
2015 Round	0.58	0.69	0.77	1	
2016 Round	0.40	0.53	0.49	0.73	1

Note: Percentage of the population who answered satisfied or rather satisfied to the following question: Would you say that you are very satisfied, rather satisfied, not very satisfied or not all satisfied with: Education to which you have access? Years between 2006 and 2017 that are not included indicate that the survey was not conducted or the question were not asked**.**
Source: OECD calculations based in the Satisfaction with Services dataset

Table 3.12. Intrawave correlation waves of measures of satisfaction with health care services between different waves of the European Quality of Life Survey

	2007	2012	2016
Wave 2 (2007)	1		
Wave 3 (2012)	0.92	1	
Wave 4 (2016)	0.84	0.90	1

Note: Percentage of the population who in a 1 (very poor quality) to 10 (high quality) scale answered 6 or more to the question. In general, how would you rate the quality of each of the following public services in [COUNTRY]? Health.
Source: OECD calculations based in the Satisfaction with Services dataset

Table 3.13. Intrawave correlation of measures of satisfaction with education services between waves of the European Quality of Life Survey

	2007	2012	2016
Wave 2 (2007)	1		
Wave 3 (2012)	0.86	1	
Wave 4 (2016)	0.77	0.92	1

Note: Percentage of the population who in a 1 (very poor quality) to 10 (high quality) scale answered 6 or more to the question. In general, how would you rate the quality of each of the following public services in [COUNTRY]? Education
Source: OECD calculations based in the Satisfaction with Services dataset

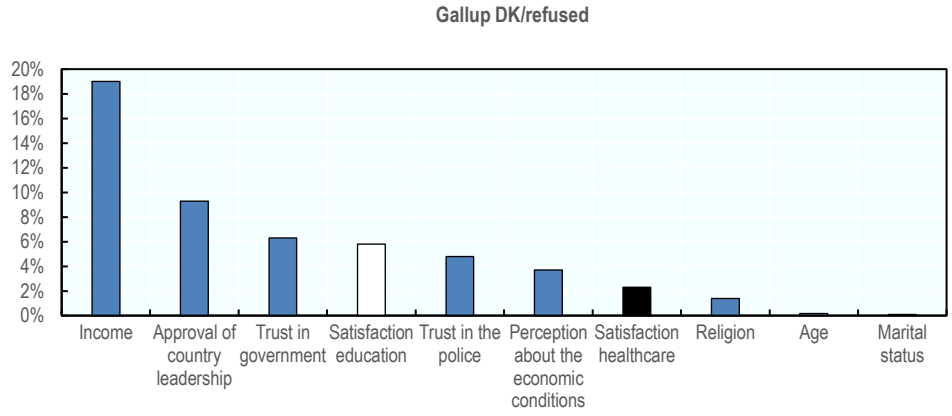
Taken together, Tables 3.12 to 3.18 support the view that measures of satisfaction with health care have a high degree of reliability. With the exception of the correlation between the Gallup World Poll and Latinobarometer (0.73) - which has fewer observations and where the questions are more distant - all other correlations are in excess of 0.8. While still respectable, evidence on the reliability of satisfaction with education measures tend to be weaker, with most correlations around 0.7. This suggests that room exists to improve such questions and strengthen the methodology for their collection.

3.4 Validity of satisfaction with education and healthcare measures

3.4.1 Face validity

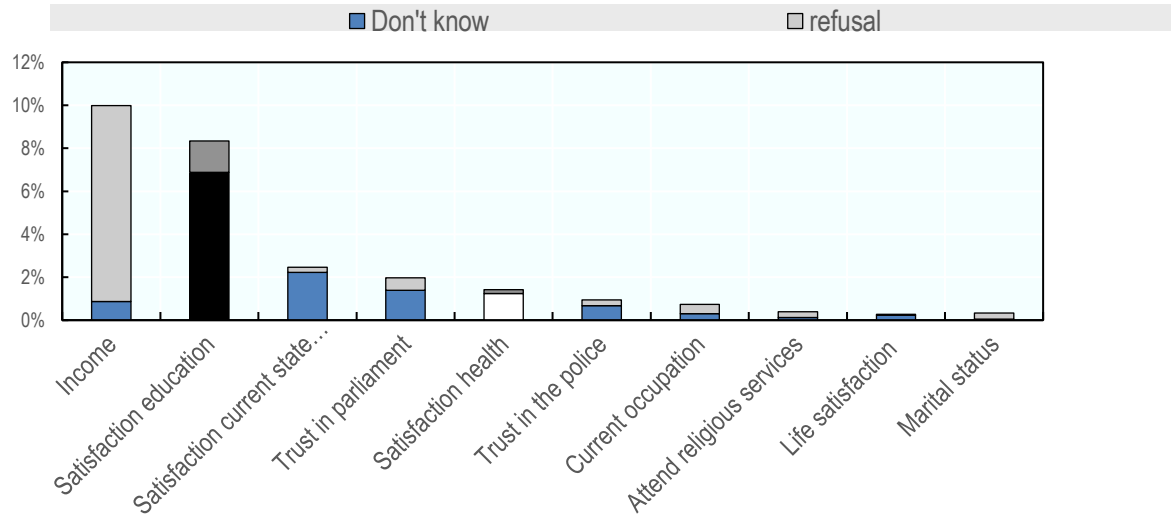
Figure 3.20 shows the item-specific non-response rates for a series of items included in the Gallup World Poll. As mentioned earlier, income questions feature relatively high non-response rates (about 19% of respondents to the Gallup World Poll didn't know or refused to provide this information). At the other end of the spectrum, demographical variables such as age or marital status have relatively low non-response rates. Satisfaction with education has non-response rate of 6%, very similar to that of trust in government (See Figure 3.20), indicating that the question may be hard to respond for some respondents, a pattern that could be explained by the fact that many respondents are far removed from the educational system. Non-response rate for satisfaction with health care is lower, the lowest among the perception questions presented in Figure 3.20. Similar results hold for the EQLS, with satisfaction with education having the second highest non-response rate (8.3%) as compared to only 2% for satisfaction with health care, indicating that most people consider having adequate information to respond this question (See Figure 3.21).

Figure 3.7. Item specific non-response in the Gallup Wold Poll



Source: OECD calculations based on the Gallup World Poll

Figure 3.8. Item-specific non-response in the EQLS, 2016



Source: OECD calculations based on the EQLS

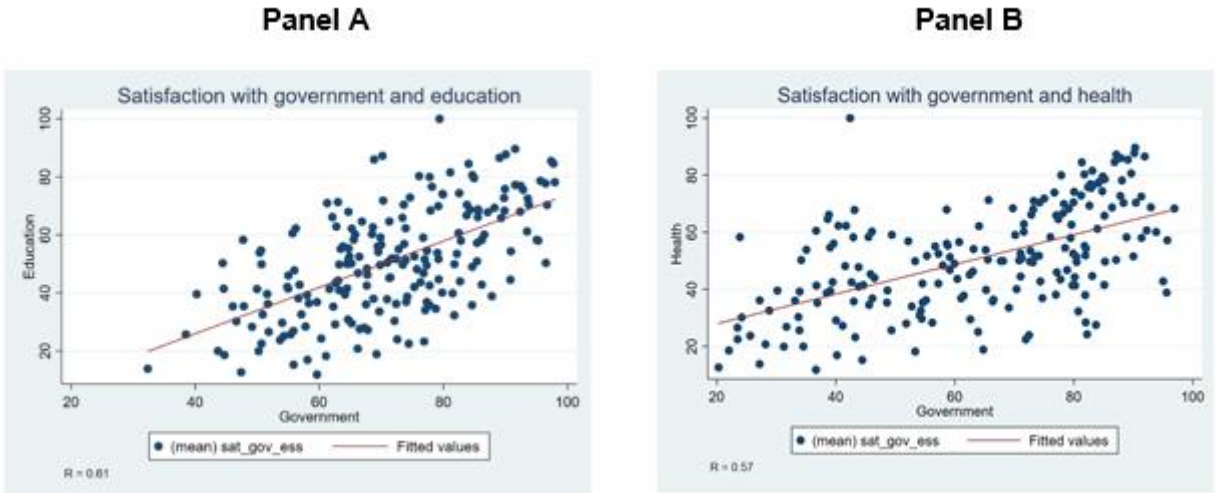
The overall picture on face validity emerging from an analysis of non-response rates is hence mixed. Satisfaction with health questions perform better than income and other perception based questions such as trust and expectations about the performance of the economy. Still, their non-response rates are above those pertaining to more objective aspects such as age and marital status, which suggests that some people lack adequate answer to provide a meaningful answer. In stark contrast, the non-response rate of questions on satisfaction with education is among the highest of perception based questions, indicating that a higher share of the population finds it difficult (or lack information) for answering this question, suggesting weaker face validity of satisfaction with education metrics.

3.4.2 Construct validity

In order to test for construct validity (i.e. degree to which a measure behaves in a way that is consistent with expectations) the three hypothesis outlined in the introductory section of this chapter will be tested in this section. These hypothesis are: 1) the existence of a halo effect or the influence that attitudes about the government or public sector could have when assessing public services in broad terms; 2) The role of experiences as a driving force shaping users' satisfaction (expectancy disconformity theory); and 3) the extent to which satisfaction is influenced by various features of the service, such affordability, proximity etc.

The first challenge for assessing the existence of the halo effect is identifying a measure capturing attitudes towards government. The ESS includes the following question "Now thinking about the [country] government, how satisfied are you with the way it is doing its job?" Figure 3.22 shows the correlation between satisfaction with government actions and, respectively, satisfaction with education (Panel A) and health care services (Panel B). In both cases the correlations are about 0.6 (slightly higher in the case of health care) indicating that satisfaction with services is weakly influenced by the perception about political leadership (i.e. how the government is doing its job).

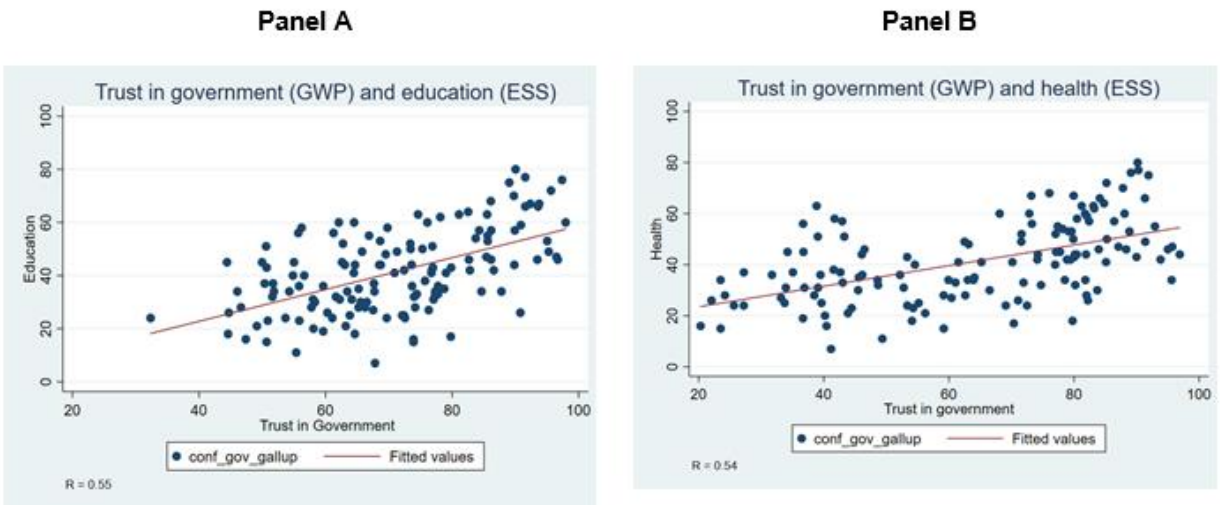
Figure 3.9. Cross-country correlations between measures of satisfaction with government and measures of satisfaction with health care and education services



Note: Percentage of the population who indicated five or more in response to the question: "Now thinking about the [country] government, how satisfied are you with the way it is doing its job?".
Percentage of the population who answered five or more to the following question "Now, using this card, please say what you think overall about the state of education in [country] nowadays?"
Source: OECD calculations based on the ESS

Another possible candidate to assess the performance of government is institutional trust. Figure 3.23 display the correlation between survey measures of people's confidence in national government from the Gallup World Poll and measures of users' satisfaction with services from the ESS. While the correlation is statistically significant, it is lower than in the case of the satisfaction with government question shown in Figure 3.22. As suggested by OECD research, survey measures of people's confidence in government have little discriminatory power (OECD/KDI 2018; OECD 2017c; González and Smith 2017). A metric of trust in civil servants would allow testing the relation between user's satisfaction and the agents directly in charge of service provision, rather than with those in charge of setting policies, however and unfortunately a comparable measure for this level of government doesn't exist.

Figure 3.10. Cross-country correlations between measures of trust in government and measures of satisfaction with health care and education services



Note: Percentage of the population who answered yes to the question; Do you have confidence in your national government
Percentage of the population who answered five or more to the following question Now, using this card, please say what you think overall about the state of education in [country] nowadays?
Source: OECD calculations based in Gallup World Poll and European Social Survey

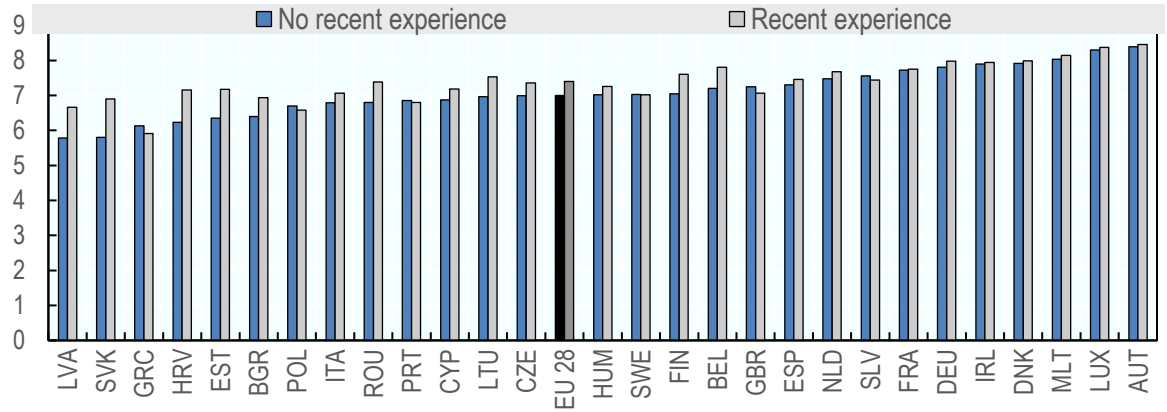
The influence that attitudes about the government could have on satisfaction measures could be reported or unconscious. Some recent U.S studies have relied on satisfaction metrics generated through Implicit Association Tests[32] (IATs) for assessing the existence of unconscious bias when users are assessing public services performance[33]. Marvel (2016) conducts several IATs to test whether respondent's assessment of the performance of the public postal service performance in the US can be biased. The hypothesis is that the socio political and cultural environment in the US, which tend to portray the public sector negatively, could overweight actual experience when users are asked to pass judgment on public services. According to Marvel's results[34], respondents' evaluations of postal service performance are weighed down by their unconscious views of the public sector, even when people are told in advance that no difference on performance exists between public and private providers. Overall, this evidence indicates that measures of people's satisfaction with health care and education services are influenced by their overall view about the government and the public sector.

The second hypothesis is that satisfaction with services will be higher/lower for respondents with a recent experience of using the service in question. Data from the European Quality of Government Index Survey (EQI) survey suggests that users with a recent experience (the past 12 months) of using different types of public services tend to report higher satisfaction than those without recent experience. While differences are small, they are statistically significant (OECD 2017a).

The same pattern holds when looking at other surveys. In a special module on satisfaction with services, the European Quality of Life Survey (EQLS) asked respondents to rate the quality of both "a general practitioner (GP), family doctor or health centre services" and that of "hospital or medical specialist services". Respondents were also asked if they or someone in their family had contact with these[35] services over the past 12 months. Figure 3.24 displays the average satisfaction (on a scale from 0-10) of the two types of health care services for people who used these services over the past 12 months and those who did not. Users' satisfaction is 0.4 percentage points higher for people with recent experience, a difference that is statistically significant; the difference is statistically significant in more than half of the European countries, although in Greece and the United Kingdom recent experience seems to result in lower satisfaction.

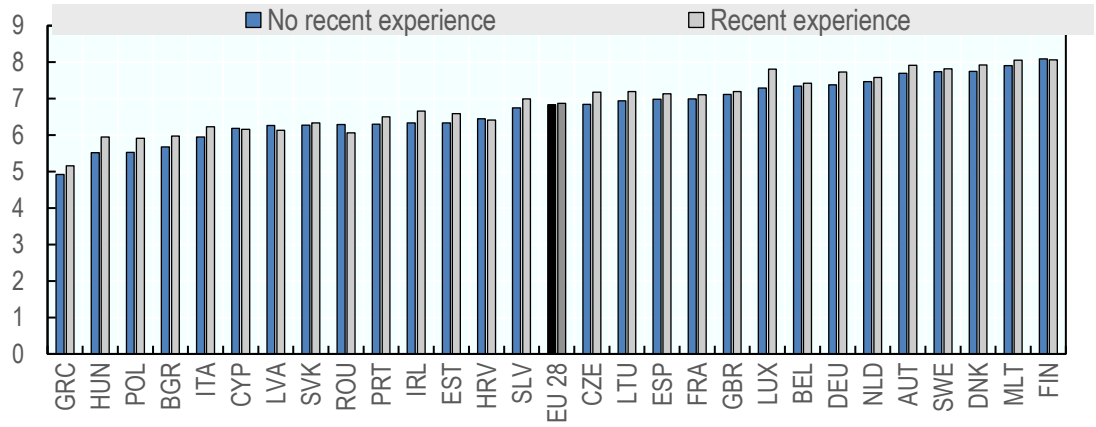
Figure 3.24 shows a similar analysis for users' satisfaction with hospitals or medical specialist. In contrast to the case of general practitioners, there are no significant differences on average in users' satisfaction between the two groups of respondents, with the difference being statistically significant in less than a third of the countries (e.g. Luxembourg, Italy and Hungary). This indicate that, when asked about satisfaction with health care in general terms people may consider simple services rather than more complex ones, or that most of the interactions take place at the GP level while visits to specialists are more rare and memory fades away as time passes. It is therefore possible to conclude that recent experience with health services influences the satisfaction with the system, at least at the more general level.

Figure 3.11. Users' satisfaction with a general practitioner, family doctor or health centre services by experience, 2016



Note: Differences across groups (no recent experience and recent experience) are assessed based on a T-test. Differences are statistically significant at 95% on average and in Latvia, the Slovak Republic, Greece, Croatia, Estonia, Bulgaria, Italy, Romania, Cyprus, Lithuania, the Czech Republic, Hungary, Finland, Belgium, the United Kingdom and the Netherlands. The group of European countries is as considered in the European Quality of Life Survey.
Source: OECD calculations based on wave 4 of the EQLS

Figure 3.12. Satisfaction with hospital or medical specialist services by experience, 2016



Note: Differences across groups (no recent experience and recent experience) are assessed based on a T-test. Differences are statistically significant at 95% in Greece, Hungary, Poland, Bulgaria, Italy, Ireland, Slovenia and Luxembourg
Source: OECD calculations based on wave 4 of the EQLS

The final aspect of construct validity refers the relation between affordability, proximity and others characteristics of the service provided, attributes that *a priori* should influence users' satisfaction. In its special module on satisfaction with services, the EQLS asked respondents to consider the last time they needed to see or be treated by a GP, family doctor or health centre, and to assess the extent to which any of the following made that difficult: a) distance to GP/doctor's office/health centre; b) delay in getting an appointment; c) waiting time ; d) cost of the visit; and e)finding time because of work and care commitments . Table 3.26 shows the cross-country correlation between users' satisfaction with healthcare and not experiencing any of the difficulties for seeing a doctor mentioned above. Overall, all correlations are statistically significant, with the only exceptions related to finding time for seeing the doctor due to other responsibilities; the highest correlations refer to distance and cost, with people don't reporting these obstacles also reporting higher satisfaction with health care services.

Table 3.14. Cross country correlation between measures of users' satisfaction with health care services and measures of different obstacles in accessing them, 2016

	Satisfaction with health	Distance to GP/doctor's office or health centre	Delay in getting an appointment	Waiting time for seeing a doctor	Cost of seeing a doctor	Finding time because of work
Satisfaction with health	1					
Distance to	0.76*	1				
Delay in getting	0.52*	0.74*	1			
Waiting time	0.60*	0.84*	0.68*	1		
Cost	0.68	0.62*	0.38	0.70*	1	
Finding time	0.40	0.50*	0.45	0.62*	0.73*	1

Source: OECD calculations based on wave 4 of the EQLS

While evidence for assessing the construct validity of measures of satisfaction with education is insufficient, in the case of health services most of the analysis is consistent with the hypothesis that users' satisfaction is influenced by various features of services presented at the introduction of this section. Yet, results need to be treated carefully as they are derived from a single survey.

Overall, the evidence presented in this section shows that broad satisfaction question asked in the Gallup World Poll may be subject to biases stemming from respondents' overall perception about the government, the lack of direct experience and the difficulty in assessing the different attributes shaping the service. AII of these factors should be considered when measuring satisfaction with health care and education services.

Table 3.27 provides a snapshot of our assessment of the accuracy of existing measures of users' satisfaction with health care and education, based on the different criteria for statistical quality considered in this paper. Overall, evidence on statistical quality is stronger for health care than for education services, with most evidence pertaining to face and construct validity rather than for convergent validity and reliability. In addition to filling the gaps for completing the reliability analysis further work could advance in standardizing the measurement of different attributes shaping satisfaction (e.g. financial and geographic access, quality of services) in household surveys for each of the services. In this regard, the survey module prepared by the EQLS and the methodology presented in the Praia Handbook may be a step in the right direction. Finally, it may be worthwhile researching how opportunities for people to participate in the design (co-design) and delivery (co-delivery) of services influence satisfaction services.

Table 3.15. Summary evaluation of the accuracy of survey measures of satisfaction with health and education service

	Face validity	Construct validity	Convergent validity	Reliability
Satisfaction with services-health	√√	√√	√	√√
Satisfaction with services-education	√	√	√	√

Note: √√ mean strong validity or reliability while √ means weak validity or reliability

4 Conclusions and statistical agenda ahead

A common criticism of governance indicators is that, regardless of their source, little evidence exists about their statistical accuracy (i.e. reliability and validity). While evidence is indeed limited, there is *a priori* no good reason why these statistics could not be generated so as to adhere to the same standards as metrics in other fields, for instance economic or environmental statistics (González et al 2017; OECD 2017a). The OECD Trust Guidelines, provide a good example of the type of evidence that would be required to perform such thorough assessment (OECD 2017c). Based on evidence from a range of non-official household surveys, this paper has summarised evidence on the statistical accuracy of survey measures of two key aspects of the responsiveness of public administrations: external political efficacy (i.e. system responsiveness) and satisfaction with services.

Political efficacy has a long, but somehow inconsistent measurement tradition. Measurement efforts have been driven by researchers and mainly relied in non-official household surveys. Still important progress has been achieved for understanding this concept. Based on this evidence, political efficacy has been identified as an indicator for measuring target 16.7.2 of the 2030 Agenda, on the *percentage of the population who believe decision making is inclusive and responsive*. The evidence in this paper shows that the so called NO-SAY question performs well enough in terms of reliability and validity for NSOs to collect these data based on their own surveys.

Still, there is room for improvement by further investigating and enhancing the performance of these metrics. For example it would be important to test whether the selection of different wording by different surveys (e.g. *political system* as finally decided in ESS and *government* in PIAAC and ISSP) has any impact on the results or if these terms are fully interchangeable (e.g. they capture the same concept). It will also be desirable to agree on a common response scale or a system of equivalences between the different scales currently being implemented in different surveys.

Testing how political efficacy questions are influenced by the closeness of elections could provide additional valuable information. While the ANES links its data collection to the electoral cycle other surveys such as ESS and ISSP follow their own internal data collection cycles, a plausible hypothesis is that people will report higher levels of political efficacy the closer the measurement is to the elections. Understanding the existence and size of this *closeness to elections* effect could shed light on what are the structural levels of the political efficacy metric and what are possible cyclical effects. In turn, there is also room for assessing the extent to which new digital technologies and additional, more frequent channels of interaction, participation and engagement between people and the public administration are affecting political efficacy levels.

Satisfaction with healthcare and education services are widely measured in non-official comparative household surveys and are commonly generated as official statistics by some NSOs. Furthermore, satisfaction with services is an SDG 16 target *16.6.2 proportion of the population satisfied with their last experience of public services.* This paper found that most surveys treat healthcare and education surveys equivalently the evidence included here shows that they are at different levels of maturity. Evidence on the accuracy of satisfaction with health metrics is stronger than for education services. In the case of health

measures there is good evidence of reliability and face validity, while evidence on construct validity is not conclusive and convergent validity cannot be tested due to the absence of proxy metrics. In turn, the accuracy of satisfaction with education measures is weaker for all reliability and validity criteria assessed in this paper.

The evidence presented here supports the view that biases towards the government or the public sector influence negatively responses on satisfaction with health and education services. Further investigating the size of these biases and if they are mitigated by more precise questions about service attributes (e.g. access, quality) could be useful for refining existing measures. Additionally, there is evidence that recent experiences with health services influence reported satisfaction. Still, even when satisfaction questions are filtered by experience there is room for improvement, by for instance investigating the effects of the intensity (i.e. a more distant or closer interaction) of experience in levels of satisfaction with services. Further work could also advance in standardizing the measurement of different attributes shaping satisfaction (e.g. financial and geographic access, quality of services) in household surveys for each of the services. In this regard, the survey module prepared by the EQLS and the methodology presented in the Praia Handbook may be a step in the right direction. Finally, it may be worthwhile researching how opportunities for people to participate in the design (co-design) and delivery (co-delivery) of services influence satisfaction services.

References

Acock, A and H. Clarke (1990), “Alternative measures of political efficacy: models and means” , Quality and Quantity, Vol 24, pp 87-105, Kluwer Academic Publishers.

Bansal, P and K. Roth (2017), “Why Companies Go Green: A Model of Ecological Responsiveness” , Academy of Management Journal, Vol 43, No 4

Barnes, L. (2013), “The political economy of working time and redistribution”, in Wren A. (ed.), The Political Economy of the Service Transition, Oxford: Oxford University Press, pp. 282-305.

Bell, S. and M. Salter Ainsworth (1972), “Infant Crying and Maternal Response Responsiveness”, Child Development, Vol 43, No 4 PP 1171-1190

Borgonovi, F. and A. Pokropek (2017), “Mind the gap: The mediating role of literacy and education in explaining disparities in external political efficacy in 22 countries”, Intelligence (2017), http://dx.doi.org/10.1016/j.intell.2017.03.006.

Bouckaert, G and S. Van de Walle (2003), “Comparing measures of citizen trust and user statisfaction as indicators of good goverance: difficulties in linking trust and satisfaction indicators” International Review of Administrative Sciences, Vol 73(4): 549-56

Bowler, S., and Donovan, T. (2002). “Democracy, institutions and attitudes about citizen influence on government”. British Journal of Political Science, 32(2), 371–390.

Campbell A. Gurin G and Miller W (1954) “The voter decides”. New YorkChamberlain, Adam. 2012. ″A Time-Series Analysis of External Efficacy.″ Public Opinion Quarterly 76: 117–30.

Carnes, N. and N. Lupu (2016), “Do voters dislike working-class candidates? Voter biases and the descriptive underrepresentation of the working class”, American Political Science Review, Vol. 110, No 4, pp. 832-844.

Carnes, N. and N. Lupu (2015), “Rethinking the comparative perspective on class and representation:Evidence from Latin America”, American Journal of Political Science, Vol. 59, pp. 1-18.

Carnes, N. (2012), “Does the numerical underrepresentation of the working class in Congress matter?”,Legislative Studies Quarterly, Vol. 37, pp. 5-34.

Clarke, H., A. Kornberg and T.J. Scotto (2010). "Accentuating the Negative?" A Political Effiacy Question-Wording-Experiment", Methodology, Vol 6(3) pp 107-117

Clarke, H and C. Alan (1989), "National Elections and Political Attitudes: The Case of Political Efficacy" British Journal of Political Science, Vol 19, pp 551-562

Craig, S and M. Maggioto (1982), " Measuring Political Efficacy", Political Methodology , Vol 8, pp 85-109.

Denemark, D and R. Niemi (2012) Political Trust, Efficacy and Engagement in Challenging Times: An Introduction, Australian Journal of Political Science, 47:1, 1-9, DOI: 10.1080/10361146.2011.643164

Dyck, J., and Lascher, E. (2009). Direct Democracy and Political Efficacy Reconsidered. Political Behavior, 31(3), 401-427. Retrieved from http://www.jstor.org/stable/40587291

Edlin, M. and S. Delamore (2018), "Ready to serve the public? The role of empathy in public service education programs", Journal of Public Affairs Education, Vol 24, No 3, pp 300-320

Eo Alenika, E (2007), "Quality of Elections, Satisfaction with Democracy and Political Trust in Africa", Afrobarometer working papers No 84

Eurofound (2019), Challenges and prospects in the EU: Quality of life and public services, Publications Office of the European Union, Luxembourg.

Feldstein, Martin. (1997). How Big Should Government Be?. National Tax Journal. 50.

Ferrín, M. (2016), "An empirical assessment of satisfaction with democracy", How Europeans View Democracy, Oxford: Oxford University Press.

Finkel, S.E. (1987), "The effects of participation on political efficacy and political support: Evidence from a West German panel", Journal of Politics, Vol. 49, No. 2, pp. 441-464.

Flavin, P. and M.J. Keane (2011), "Life satisfaction and political participation: Evidence from the United States", Journal of Happiness Studies, Vol. 13, No. 1, pp. 63-76.

France Stratégie (2016), Lignes de Faille. Une Société à Réunifier, Paris, www.strategie.gouv.fr/sites/strategie.gouv.fr/files/atoms/files/dp_-_lignes_de_faille_0.pdf (accessed on 17 March 2017).

Gilens, M. (2005), "Inequality and democratic responsiveness", Public Opinion Quarterly, Vol. 69, No. 5, Special Issue 2005, pp. 778-796.

Hetherington, Marc J. (1998). "The Political Relevance of Political Trust." American Political Science Review 92: 791–808.

González , S., L Fleischer and M. Mira d'Ercole (2017), "Governance Statistics in OECD countries and beyond: What exists , and what would be required to assess their quality", *OECD Statistics Working Papers,* 2017/03, OECD Publishing, Paris. Htttp//dx.doi.org/10.1787/c0d45b5e-en

González, S. and C. Smith (2017), "The accuracy of measures of institutional trust in household surveys: Evidence from the oecd trust database", OECD Statistics Working Papers,2017/11, OECD Publishing, Paris. http://dx.doi.org/10.1787/d839bd50-en

Geurkink, B., Zaslove, A., Sluiter, R., and Jacobs, K. (2020). Populist Attitudes, Political Trust, and External Political Efficacy: Old Wine in New Bottles? Political Studies, 68(1), 247–267. https://doi.org/10.1177/0032321719842768

Hobolt, S.V. and R. Klemmensen (2008), "Government Responsiveness and Political Competition in Comparative Perspective", Comparative Political Studies, Vol 41, No 3, pp 309-337, Sage publications.

IDEA (International Institute for Democracy and Electoral Assistance) (2016), Voter Turnout Trends around theWorld, IDEA Publishing, Stockholm.

Ikeda, K., T Kobayahi and M. Haashimoto (2008), " Does political participation make a difference? The relationship between political choice, civic engagement and political efficacy". Electoral Studies,

Ivanyna, M. and A. Shaw (2018), "Ranking Countries for Good Governance Using Public Opinion Surveys", The Palgrave Handbook of Indicators in Global Governance". D.V Malito et al. (eds), Springer Nature.

Jacobsen, R., J Snyder and Saultz, A (2014), "Informing or Shaping Public Opinion? The Influence of School Accountability Data Format on Public Perceptions of School Quality" American Journal of Education, Vol. 121, No. 1 (November 2014), pp. 1-27

James, O. (2009), "Evaluating the expectations disconfirmation and expectations anchoring approaches to citizen satisfaction with local public services". Journal of Public Administration Research and Theory, Vol 19, No 1 pp. 107-23

Jakobsen, M and R. Jensen (2015) Common Method Bias in Public Management Studies, International Public Management Journal, 18:1, 3-30, DOI:10.1080/10967494.2014.997906

Karp, J. and Banducci, S. (2008). Political Efficacy and Participation in Twenty-Seven Democracies: How Electoral Systems Shape Political Behaviour. British Journal of Political Science, 38(2), 311-334. Retrieved from http://www.jstor.org/stable/27568347

Lane (1959), Political Life: Why and How People Get Involved in Politics, Free Press, New York.

Marina, B. (2015), "Managing Diversity in Communities, Workplace and Societies" ,Diversity and Public Administration, Mitchell Rice ed, Routledge, New York

Marvel, J.D. (2016), "Unconscious bias in citizens evaluations of public sector performance", Journal of Public Administration Research and Theory , Vol 26, No 1, pp. 143-158

Mcevoy, C. (2016), "The role of political efficacy on public opinion in the European Union", Journal of Common Market Studies, Vol. 54, No. 5, pp. 1159-1174.

McPherson, J., Welch, S., and Clark, C. (1977). The Stability and Reliability of Political Efficacy: Using Path Analysis to Test Alternative Models. *American Political Science Review, 71*(2), 509-521. doi:10.1017/S0003055400267427

Miller, A and O. Lishtaug (1990), "Political parties and confidence in government: a comparison of Norway, Sweden and the United States" British Journal of Political Science, Vol 20, No 3, pp 357-386

Miller, W., A. Miller and E. Schneider (1980), "American National Election Studies Data Sourcebook, 1952-1978" Harvard University Press, Cambride MA

Moore, J.C., L Stitson and E. Welniak (2000), "Income measurement error in surveys: A review", *Journal of Official Statistics*, Vol. 16.

Niemi, R., S. Craig and F. Mattei (1991), "Measuring Internal Political Efficacy in the 1988 National Election Study" American Political Science Review, Vol 85, No 4,

Young Mok, K.,J. Oliver and G. Van Ryzin (2017), "Expectations of and Satisfaction with Public Services", Experiments in Public Management Research, Cambridge University Press

OECD (2019a), Government at a Glance 2019, OECD Publishing, Paris, https://doi.org/10.1787/8ccf5c38-en.

OECD (2019b), Measuring What Matters: The Patient-Reported Indicator Surveys, Status Report, OECD Publishing, Paris.

OECD (2019c), Health at a Glance 2019: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/4dd50c09-en.

OECD (2019d), Education at a Glance 2019: OECD Indicators, OECD Publishing, Paris, https://doi.org/10.1787/f8d7880d-en

OECD (2017a), How's Life? 2017: MeasuringWell-being, Governance and Well-being, OECD Publishing, Paris.http://dx.doi.org/10.1787/how_life-2017-en

OECD (2017b), OECD (2017), Government at a Glance 2017, OECD Publishing, Parishttp://dx.doi.org/10.1787/gov_glance-2017-en

OECD (2017c), OECD Guidelines on Measuring Trust, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264278219-en

OECD (2016), Government at a Glance: Latin America and the Caribbean 2017, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264265554-en

OECD (2015), How's Life? 2015: Measuring Well-being, OECD Publishing, Paris, http://dx.doi.org/10.1787/how_life-2015-en.

OECD (2014f), Women, Government and Policy Making in OECD Countries: Fostering Diversity for Inclusive Growth, OECD, Paris, http://dx.doi.org/10.1787/9789264210745-en.

OECD/KDI (2018), Understanding the Drivers of Trust in Government Institutions in Korea, OECD Publishing, Paris.https://doi.org/10.1787/9789264308992-en

OECD (2011), Public Servants as Partners for Growth: Toward a Stronger, Leaner and More Equitable Workforce, OECD, Paris, http://dx.doi.org/10.1787/9789264166707-en.

OECD (2008), "Quality Framework of OECD Statistics", OECD Publishing, Paris

Oliver, R. (2010), Satisfaction: A Behavioural perspective on the consumer, Routledge, New York.

Pitts, D.W. and L.R.Wise (2010), "Workforce Diversity in the New Millennium: Prospects for Research",Review of Public Personnel Administration, Vol. 30, No. 1, pp. 44-69.

Pollock, Philip H., III. (1983). "The Participatory Consequences of Internal and External Political Efficacy: A Research Note." The Western Political Quarterly 36: 400–09.

Piani A et al (2000), "Expression of MHC class II molecules contributes to lipopolysaccharide responsiveness" Division of Immunology, Zurich University Children's Hospital, https://doi.org/10.1002/1521-4141(200011)30:11<3140::AID-IMMU3140>3.0.CO;2-O

Putnam, R (2005), Bowling Alone, the Collapse and Revival of American Community, Simon and Schuster, New York

Pratima, B and K. Roth (2017), "Why companies Go Green: a Model of Ecological Responsiveness" Academy of Management Journal, Vol 43 No 4, https://doi.org/10.5465/1556363

Roch, C.H and T. Poisner (2006), " Citizens, accountability, and service satisfaction the influence of expectations" Urban Affairs Review, Vol 41, No 3, pp 292-308

Saris, W., and M. Torcal (2009), "Alternative measurement procedures and models Political Efficacy" Research and Expertise Centre for Survey Methodology (RECSM), RECSM Working Paper Number 3, Universitat Pompeu Fabra, Barcelona

Selden, S.C. and F. Selden (2001), "Rethinking diversity in public organizations for the 21st century:Moving toward a multicultural model" Administration and Society, Vol. 33, No.3, pp 303-329

Smith, D. A., and Tolbert, C. J. (2004). Educated by initiative: The effects of direct democracy on citizens and political organizations in the American States. Ann Arbor: The University of Michigan Press.

UNDP (2017) " A review and testing of available indicators for SDG 16.7.2:Proportion of population who believe that decision-making is inclusive and responsive", UNDP Oslo Governance Centre

UNDP (2019a) " SDG Indicator 16.7.2 Proportion of the population who believe decision-making is inclusive and responsive, by sex, age, disability and population group; Methodology Development Narrative", UNDP Oslo Governance Centre

UNDP (2019b) " SDG Indicator 16.6.2 Proportion of the population satisfied with their last experience of public services; Methodology Development Narrative ", UNDP Oslo Governance Centre

Uslaner, E. M., & Brown, M. (2005). Inequality, Trust, and Civic Engagement. American Politics Research, 33(6), 868–894. https://doi.org/10.1177/1532673X04271903

Van de Walle, S (2018). "Explaining citizen satisfaction and dissatisfaction with public services", The Palgrave Handbook of Public Administration and Management in Europe. Palgrave Macmillan. London.

Van de Walle, S. and G. Bouckaert (2013) “Public Service performance and trust in government: the problem of causality”, International Journal of Public Administration , Vol 26, No 8-9, pp 891-913

Van Ryzin, G (2006), “Testing the Expectancy Disconfirmation Model of Citizen Satisfaction with Local Government” *Journal of Public Administration Research and Theory*, Volume 16, Issue 4, October 2006, Pages 599–611, https://doi.org/10.1093/jopart/mui058

Van Ryzin, G (2004), G “Expectations, Performance, and Citizen Satisfaction with Urban Services”, Journal of Policy Analysis and Management, Vol 23, No 3, 433-448

White, H and M. Rice (2015), “The Multiple Dimensions of Diversity and Culture” , Diversity and Public Administration, Mitchell Rice ed, Routledge, New York

Yang, K and Holzer, M (2006), “The Performance-Trust Link: Implications for Performance Measurement”, Public Administration Review, Vol 66, pp 114-126

Annex A. Table external political efficacy questions in selected survey modules.

Survey	Questions	Response scale	Time period	Institution responsible
ANES (since 1988)	Do you agree or disagree with the following statements? Sometimes politics and government seem so complicated that a person like me can't really understand what's going on **People like me don't have a say in what the government does** I don't think public officials care much what people like me think	Agree/Disagree/Neither agree nor disagree	1952-ongoing	Survey Research Centre (SRC) and Centre for Political Studies (CPS) at University of Michigan
Political Support in Canada Project (PSC)	Do you agree or disagree with the following statements? Generally, those elected to parliament in Ottawa soon lose touch with people? I don't think that the federal government cares much about what people like me think? Federal political parties are only interested in people's votes but not in their opinions **People like me don't have a say about what the federal government does?** Sometimes, politics and government in Ottawa seem so complicated that a	Agree/Disagree	1983-1993	Inter-University Consortium for Political and Social Research

	person like me can't really understand what's going on.			
European Social survey round 7	**How much would you say the political system allows people like you to have a say in what the government does?** How much would you say the political system in [country] allows people like you to have an influence in politics? How much would you say that politicians care what people like you think? How easy do you personally find it to take part in politics?	From 0 to 10 (with anchors on the extremes)	2014 (testing phase)	European Research Infrastructure Consortium (ERIC)
European Social round 8	**How much would you say the political system allows people like you to have a say in what the government does?** And how much would you say the political system in [country] allows people like you to have an influence in politics?	Not at all/very little/some/a lot/a great deal	2016-ongoing (final inclusion)	European Research Infrastructure Consortium (ERIC)
National Survey on Political Culture and Citizens practices	Do you agree or disagree with the following statements? **People like me don't have a say about what the government does?**	A lot/little/none	2001, 2003,	Ministry of Interior/ INEGI
National Survey on Political Culture and Citizens practices	Do you agree or disagree with the following statements? People in government don't care about what people like me think? **People like me don't have a say about what the government does?** Voting is the only way people like me could have influence on	Agree/Disagree/Neither agree nor disagree	2008	Ministry of interior/INEGI

	what the government does?			
National Survey on Political Culture and Citizens practices/Mexico	How much do you think that citizens could influence government decisions? (round 5)	A lot/little/none	2005, 2012	Ministry of Interior/Private polling companies
PIAAC	To what extent do you agree or disagree with the following statement? People like me don't have any say in what the government does.	From 1 (strongly agree) to 5 (strongly disagree)	Around 2012	OECD
Strategy for the Harmonisation of Statistics in Africa (SHaSA)	How often do you think the following listen to people like you? A. Members of Parliament/National Assembly B. Local elected officials/councilors C. Traditional leaders	Never (1), Sometimes (2),Often (3) and Always (4)	2001-ongoing	National Statistical Offices in African and Latin American countries

Annex B. Questions on satisfaction with health and education services in selected survey modules.

Survey	Questions	Response scale	Time period	Institution responsible
World Poll	• In the city or area where you live, are you satisfied or dissatisfied with the education system or the schools? • In the city or area where you live, are you satisfied or dissatisfied with the availability and quality of health care?	Satisfied/Dissatisfied	2006-2017	Gallup
European Quality of Life Surveys	In general, how would you rate the quality of each of the following public services in [COUNTRY]? • Health • Education • Public transportation • Child care services • Long-term care services • Social municipal housing	From 1 (very poor quality) to 10 (high quality)	2003, 2007, 2011, 2016	European Foundation for the Improvement of Living and Working Conditions
European Social Survey	• Card 12. Now, using this card, please say what you think overall about the state of education in [COUNTRY] nowadays? • Still Card 12. Still using this card,	From 0 (extremely bad) to 10 (extremely good)	2002, 2004, 2006, 2008, 2010, 2012, 2014, 2016	

	please say what you think overall about the state of health services in [country] nowadays?			
European Quality of Government Index	• Qu1: Have you or any of your immediate family been enrolled or employed in the public school system in your area in the past 12 months? • Qu2: In the past 12 months have you or anyone in your immediate family used public health care services in your area? • Qu3: Have you or anyone in your immediate family had any recent contact (positive or negative) with the security or police forces in your area in the past 12 months? • Qu4: How would you rate the quality of public education in your area? • Qu5: How would you rate the quality of the public health care system in your area? • Qu6: How would you rate the quality of the police force in your area?	• Qu1, Qu2, Qu3: 1 (yes), 2 (no), 99 (Don't know/Refused) • Qu4, Qu5, Qu6: from 0 (very poor) to 10 (excellent quality)	2010, 2013, 2017	University of Gothenburg
Latinobarometer	*[Translated from Spanish]* Would you say that you are very satisfied, rather satisfied, not very satisfied or not all satisfied with: • Health care to which you have access • Education to which you have access Now I would like to ask you	• 1 (very satifisfied), 2 (rather satisfied), 3 (not very satisfied), 4 (not at all satisfied), 0 (don't know/no response) • 1 (yes), 2 (no), 8 (don't know), 0 (no response)	1995-2018	

	about the experience you have had with regard to public services. • In the last 12 months, have you had contact with a public school? • In the last 12 months, have you had contact with a public clinic or hospital?			
Strategy for the Harmonisation of Statistics in Africa (SHaSA)	Have you used the following services institutions during the last 12 months? Do you trust them D. Public health system E. Public education system	Yes/No	2001-ongoing	National Statistical Offices in African and Latin American countries

Annex C. Results of the confirmatory factor analysis political efficacy modules

```
Measurement:  psppsgva psppipl ptcpplt etapapl

Exogenous variables

Latent:       EFFICACY

Fitting target model:

Iteration 0:   log likelihood =  -324828.8
Iteration 1:   log likelihood = -324814.68
Iteration 2:   log likelihood = -324814.65

Structural equation model                       Number of obs     =     37,673
Estimation method  = ml
Log likelihood     = -324814.65

 ( 1)  [psppsgva]EFFICACY = 1
```

Standardized	Coef.	OIM Std. Err.	z	P>|z|	[95% Conf.	Interval]
Measurement						
psppsgva						
EFFICACY	.7720109	.0026756	288.54	0.000	.7667669	.7772549
_cons	1.381045	.0072012	191.78	0.000	1.366931	1.395159
psppipl						
EFFICACY	.8786334	.0022038	398.69	0.000	.874314	.8829527
_cons	1.339141	.0070954	188.73	0.000	1.325234	1.353047
ptcpplt						
EFFICACY	.7438712	.0029073	255.86	0.000	.738173	.7495694
_cons	1.339947	.0070974	188.79	0.000	1.326036	1.353858
etapapl						
EFFICACY	.5999968	.0037309	160.82	0.000	.5926845	.6073092
_cons	1.329131	.0070704	187.99	0.000	1.315273	1.342989
var(e.psppsgva)	.4039991	.0041311			.3959828	.4121777
var(e.psppipl)	.2280034	.0038727			.2205381	.2357215
var(e.ptcpplt)	.4466556	.0043253			.4382581	.4552141
var(e.etapapl)	.6400038	.004477			.6312889	.648839
var(EFFICACY)	1	.			.	.

```
LR test of model vs. saturated: chi2(2)   =    440.44, Prob > chi2 = 0.0000
```

Fit statistic	Value	Description
Likelihood ratio		
chi2_ms(2)	440.441	model vs. saturated
p > chi2	0.000	
chi2_bs(6)	60741.490	baseline vs. saturated
p > chi2	0.000	
Population error		
RMSEA	0.076	Root mean squared error of approximation
90% CI, lower bound	0.070	
upper bound	0.082	
pclose	0.000	Probability RMSEA <= 0.05
Information criteria		
AIC	649653.306	Akaike's information criterion
BIC	649755.746	Bayesian information criterion
Baseline comparison		
CFI	0.993	Comparative fit index
TLI	0.978	Tucker-Lewis index
Size of residuals		
SRMR	0.016	Standardized root mean squared residual
CD	0.869	Coefficient of determination

[1] Cambridge Dictionary

[2] The taxonomy used in this paper is not comprehensive. Responsiveness as a dimension of public governance comprises more than the two specific aspects considered in this paper. Different stimuli to institutions will require different responses from institutions. For example, an outbreak of an epidemics.

[3] The Sustainable Development Goals (SDGs) are a collection of 17 global goals designed to be a "blueprint to achieve a better and more sustainable future for all". The SDGs, set in 2015 by the United Nations General Assembly and intended to be achieved by the year 2030, are part of UN Resolution 70/1. Goal 16 in particular seeks to "promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build, effective accountable and inclusive institutions at all levels"

[4] Research on political efficacy often distinguishes between internal efficacy (i.e. having the personal competence to participate in politics) and external efficacy (i.e. a belief in the responsiveness of public institutions and government officials to citizens' demands). Only the second concept is discussed in this paper, as it relates to people's perceptions of government responsiveness.

[5] Before 1982 the Centre For Political Studies (CPS) through its American National Election Studies (ANES) measured political efficacy, using a agreed/disagreed scale, by asking survey respondents if they "agreed" or "disagreed" with several statements: 1) People like me don't have a say about what the government does 2) Voting is the only way that people like me can have any say about how the government run things 3) Sometimes politics and government seem so complicated that a person like me can't really understand what's going on 4) I don't think public officials care much about what people like me think 5) Generally speaking, those we elect to congress in Washington lose touch with the people pretty quickly 6) Parties are only interested in people's votes but not in their opinions.

6 The American National Election Studies (ANES) are academically-run national surveys of voters in the United States, conducted before and after every presidential election. The ANES was formally established by a National Science Foundation grant in 1977; however, the data are a continuation of studies going back to 1948. The study has been based at the University of Michigan since its origin and, since 2005, has been run in partnership with Stanford University.

7 The four new internal efficacy items tested in 1997-1998 are the following: 1) I consider myself to be well qualified to participate in politics 2) I feel that I have a pretty good understanding of the important political issues facing our country 3) I feel that I can do as good a job in public office as most people 4) I think that I am better informed about politics and government than most people.

8 External political efficacy was measured in PIAAC by presenting respondents with the following statement: "People like me do not have a say in what the government does" to which respondents could answer on a 5 point Likert scale ranging from "strongly agree", "agree", "neither agree nor disagree", "disagree" to "strongly disagree".

9 The two questions on internal political efficacy that will be added to the core module of the ESS are the following: A) How able do you think you are to take an active role in a group involved with political issues? 1(not at all able), 2 (a little able), 3(quite able), 4(very able) 5 (completely able) and B) And how confident are you in your own ability to participate in politics 1(Not at all confident), 2 (A little confident), 3(Quite confident), 4(Very confident) 5 (Completely confident)

10 The 2014 testing round of the ESS included a 0-10 scale

11 According to these researchers the use of questions as opposed to statements improves the reliability of the tested model and provides a better representation of the system responsiveness concept.

12 The original question in Spanish is formulated in the following way: Para cada una de estas frases, por favor dígame, ¿está usted de acuerdo o en desacuerdo? La gente como yo tiene influencia sobre lo que hace el gobierno. a) muy de acuerdo b) algo de acuerdo c) ni de acuerdo ni desacuerdo d) algo en desacuerdo e) muy en desacuerdo

13 The original question in Spanish is formulated in the following way: Dígame si usted está de acuerdo o no con cada una de las siguientes frases: Las personas como usted no tienen nada que opinar sobre lo que hace el gobierno. Sí esta de acuerdo, ni de acuerdo ni en desacuerdo, esta en desacuerdo.

14 According to the expectancy-disconfirmation model satisfaction should be interpreted as a conjunction of experience with knowledge about prior expectations. In turn, those expectations act as predictors of future performance, or an anticipation of what will follow, that exists prior to the service experienced. Perceived service performance is then evaluated in comparison to the original expectations, leading to a confirmation of or disconfirmation of expectations, which influences satisfaction.

15 A distinction is made between a vertical and a horizontal dimension used to understand satisfaction. The vertical dimension refers to satisfaction with a single event while the horizontal refers to an aggregation over many situations and time periods.

16 The countries are Cameroon, Germany, Kenya, Mexico, Viet Nam, the Philippines, South Africa, Latvia, Norway, Tunisia, New Zealand,

17 Different surveys include different services. For example in addition to the Health and Education the Gallup World Poll also asks about satisfaction with public transportation, air quality and water quality. In addition to health and education the European Quality of Life Survey also collects data on long term care (e.g. nursing etc). In several waves the Latinobarometer has included questions about satisfaction with administrative services (e.g. obtaining passports, licences, permits etc)

18 Justice Services are the focus of another chapter of the handbook.

19 Commonly data are collected annually however in some years the questionnaire hasn't been fielded (i.e. 1999 and 2012)

20 The exact wording of the question is the following: Would you say that you are very satisfied (1), fairly satisfied (2), not very satisfied (3) or not at all satisfied (4), with the way it work: public hospitals, public education, police, judicial system, the place where you obtain an ID card.

21 This change is probably motivated by the existence of dual education and health systems in LAC countries as well as an important share of the population, the wealthier, opting out from the public system (OECD 2016).

22 The Serving Citizens Framework itself provides a narrow definition of responsiveness as the extent to which services provided respond to citizen's needs.

23 Contingent validity is a fourth criterion of validity that applies where validity can be directly assessed. For example, a measure of taxes paid has contingent validity if it is compiled from the complete set of tax payments done by a person. This measure is necessarily valid contingent on using a data source that directly measures the concept of interest

24 In survey research, the standard measure of reliability is test-retest reliability where the same measurement item is administered to the same person after a delay of some period. This may be later on in the same survey, or it may involve the respondent being re-surveyed after a fixed period of time

25 Generally, the technique allows the separation of construct stability and indicator reliability in over time analysis; specification of non-random error is also possible.

26 Acquisence bias refers to the tendency by respondents to voice "unthinking" agreement with survey statements as part of efforts to please the interviewer or complete the task faster.

27 Most research on the field of institutional trust has relied on questions about trust/confidence in national government. Recent research has demonstrated that this question does not discriminate between political and institutional aspects of trust

28 Unfortunately no data on trust in the civil service is available on the ESS and therefore is not possible to assess the relation with aspects of institutional trust as differentiated from political trust.

29 The liberal variant of democracy includes an electoral component (the characteristics of the electoral process) and a liberal component (existence of civil liberties and a functioning public sphere)

30 Wave 6 of the European Social Survey included a special rotating module on citizens' valuations of different elements of democracy including questions on possibilities for direct participation. It is expected that the special rotating module on citizens' valuation of democracy will be fielded again in 2020 allowing for comparison between political efficacy and possibilities for direct participation.

31 Chi square 440.44 prob>chi2=0.0000; RMSEA=0.076; CFI=0.993;SRMR=0.016

32 IATs are a psychometric technique used to test respondent attitudes where issues of social desirability may make them unwilling to respond honestly, or in areas that are difficult to measure through explicit self-reporting due to lack of awareness. In this case Implicit Association Test (IAT) are used to produce an experimental measure of trust in institutions. It asks respondents to rapidly sort relevant words to the left and right hand sides of the computer screen. The IAT relies on the idea that a person will react more quickly when the concept and the evaluation that she makes of this concept are congruent in her subconscious

33 The metric generated by these tests are IAT scores that capture th difference in sorting times for matching (USPS/slow and Fedex/fast) and unmatching categories (USPS/fast and Fedex/Slow) on a computer screen

34 The author has several designs of the IAT tests. In all designs respondents are split between a treatment and a control group and exposed to different types of information/advertisement about the performance of postal services.

35 In addition respondents are also asked if they have had contact with emergency healthcare; ordering prescriptions online or by telephone; medical consultation online or by telephone.