OECD Guidelines on Measuring Trust

OECD

OECD Better Life Initiative

OECD Guidelines on Measuring Trust

OECD
BETTER POLICIES FOR BETTER LIVES

This work is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of OECD member countries or those of the European Union.

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.

Please cite this publication as:
OECD (2017), *OECD Guidelines on Measuring Trust*, OECD Publishing, Paris.
http://dx.doi.org/10.1787/9789264278219-en

ISBN 978-92-64-27820-2 (print)
ISBN 978-92-64-27821-9 (PDF)
ISBN 978-92-64-27822-6 (epub)

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Corrigenda to OECD publications may be found on line at: *www.oecd.org/about/publishing/corrigenda.htm*.

© OECD 2017

You can copy, download or print OECD content for your own use, and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgment of the source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to *rights@oecd.org*. Requests for permission to photocopy portions of this material for public or commercial use shall be addressed directly to the Copyright Clearance Center (CCC) at *info@copyright.com* or the Centre français d'exploitation du droit de copie (CFC) at *contact@cfcopies.com*.

Foreword

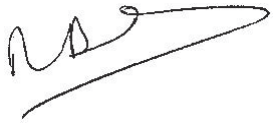
Understanding and improving well-being requires solid evidence that can inform policymakers and citizens where, when, and for whom life is getting better. Since the Stiglitz-Sen-Fitoussi Commission highlighted the need to complement GDP with better measures of social, economic and environmental outcomes in 2009, the statistical community has made remarkable progress towards developing and producing such measures and regularly monitoring human well-being. Nevertheless, certain topics have not yet received the attention their importance for society's progress might warrant. Trust is one of these topics, and these Guidelines on Measuring Trust, *prepared under the umbrella of the OECD Better Life Initiative launched in 2011, represent an important step towards improving and expanding the system of well-being statistics further.*

The issue of trust, or lack of it, has mostly made headlines in the aftermath of the 2008 financial crisis. Indeed, people's trust in their public institutions fell sharply in most OECD countries and it has not since fully recovered to its pre-crisis levels. Yet, only a society where people cooperate with and express solidarity for one another, and where public institutions act competently and are accessible to all citizens enables a high quality of life for all. Trust in other people and trust in institutions are essential ingredients for social and economic progress while a prospering society, in turn, is one in which trust can blossom.

It is therefore no surprise that several recent policy initiatives have stressed the need for better measures of trust. For instance, the OECD Trust Strategy was initiated during the 2013 OECD Ministerial Council meeting on Jobs, Equality and Trust to provide guidance, including methodological and measurement advice, to OECD governments on how to restore trust in public institutions. As well, Goal 16 of the Sustainable Development Goals (SDGs) endorsed in 2015 by all UN member countries ("Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels") focuses on trust and governance. And the UN Statistical Commission has set up a dedicated group (the Praia Group) to develop a handbook on governance statistics to inform and monitor the SDG targets under Goal 16.

With some notable exceptions, the measurement of trust does not have a long tradition, particularly within official statistics, and official measures that exist are not always collected in a regular and internationally comparable manner. These Guidelines aim at contributing to filling this gap. Their main objective is to support data producers in their own initiatives to measuring trust. In particular, the Guidelines will provide direct inputs to the UN Praia Group. They synthesize what we currently know about good practice on how trust can, and should, be measured. This knowledge might change as the evidence base on trust develops. However, especially so for measures of interpersonal trust, there is already today good evidence that the suggested survey questions produce valid data and are ready to be included in official surveys.

It is my hope that these Guidelines will contribute to a step change in the quality and availability of trust data going forward.



Martine Durand
OECD Chief Statistician
Director of the OECD Statistics Directorate

Acknowledgements

This report is published under the responsibility of the Secretary-General of the OECD. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the Organisation or of the governments of its member countries.

These Guidelines were produced as part of the work programme of the OECD Statistics Committee and the Public Governance Committee, whose delegates have reviewed the report. The report was prepared by Conal Smith and Lara Fleischer from the OECD Statistics Directorate. An advisory group composed of representatives from national statistical offices of OECD member countries (Adrian Franco, INEGI, Mexico; Sophie Pontieux, INSEE, France; Dawn Snape and Veronique Siegler, ONS, UK; Fiona Carnes and Joanne Baker, ABS, Australia), and researchers and policy makers (Prof. Soonhee Kim and Prof. Dong-Young Kim, Korea Development Institute; Prof. Jacob S. Hacker, Yale University; Prof. Yann Algan, Paris Institute of Political Studies; Dr. Monica Ferrin, University of Turin; Jacob Saeger, UK Cabinet Strategy Office; Prof. Marc Hetherington, Vanderbilt University; Prof. Eugene Kandel, Hebrew University of Jerusalem and former Chairman of the Israeli National Economic Council) provided very valuable comments and advice on the drafting of the report. The report also benefited from the comments of Martine Durand, Marco Mira d'Ercole, Carrie Exton and Fabrice Murtin (OECD Statistics Directorate), as well as Rolf Alter, Zsuzsanna Lonti, Paloma Baena Olabe and Santiago Gonzalez (OECD Directorate for Public Governance). Patrick Hamm, Anne-Lise Faron and Virginie Elgrably ensured the final editing of the document.

Table of contents

Executive summary 11

Overview and recommendations 15
1. Concept and validity 16
2. Methodological considerations in the measurement of trust 17
3. Measuring trust 19
4. Output and analysis of trust data 23
5. Next steps 24

Chapter 1. **Introduction** 25
1.1. Introduction 26
1.2. Motivation 26
1.3. The Guidelines for measuring trust 29
1.4. Conclusion 32
Notes 32
References 32

Chapter 2. **Concept and validity** 35
2.1. Introduction 36
2.2. A conceptual framework for measuring trust 36
2.3. Approaches to measuring trust 43
2.4. The statistical quality of trust measures 47
2.5. Conclusion 78
Notes 79
References 80

Chapter 3. **Methodological considerations** 85
3.1. Introduction 86
3.2. Measurement error 86
3.3. Question wording 87
3.4. Response formats 94
3.5. Survey context 98
3.6. Survey mode 101
3.7. Response styles and the cultural context 104
3.8. Conclusion 107
Notes 109
References 110

Chapter 4. **Measuring trust** 115
4.1. Introduction 116
4.2. What to measure? Planning the measurement of trust. 117
4.3. Survey and sample design 124
4.4. Questionnaire design 132
4.5. Survey implementation 147
4.6. Conclusion 148
Notes. 152
References 153

Chapter 5. **Output and analysis of trust data** 155
5.1. Introduction 156
5.2. Reporting trust data. 156
5.3. Interpreting trust data. 165
5.4. Analysing trust data 175
5.5. Conclusion 181
Notes. 182
References 183

Annex 1. **Illustrative examples of trust measures**. 186

Annex 2. **Question modules** 196

Tables

2.1. A framework for multiple trust relationships 39
2.2. The dimensionality of trust measures in the World Values Survey 41
2.3. Surveys included in the OECD Trust Database and their key characteristics . . . 54
2.5. Correlation in trust in the judicial system across surveys. 57
2.4. Correlation in trust in the national government across surveys 57
2.6. Intra-wave correlation for generalised trust, based on the European Social Survey. 59
3.1. Overview of response biases and heuristics 87
3.2. Comparison of interpersonal trust questions by gender and age groups 90
3.3. Comparison of confidence in institutions versus confidence in institutions to act in the national interest. 91
3.4. Comparison of confidence in institutions versus trust in institutions 93
4.1. The OECD Trust Framework 142
5.1. Suggested structure of statistical releases of trust data 157
5.2. Interpersonal trust by population group in Canada in selected years. 164
5.3. Magnitude of differences for measures of interpersonal trust 166
5.4. Magnitude of differences for measures of trust in the police 167
5.5. Magnitude of differences for measures of trust in the judicial system 168
5.6. Magnitude of differences for measures of trust in government 169

Figures

2.1. Correlation between measures of generalised trust across surveys 59
2.2. Intra-wave correlation for generalised trust based on the European Quality of Life Survey 60

2.3. Item specific non-response in the European Social Survey, 2002-14 61
2.4. Item-specific non-response rates in the Gallup World Poll, 2006-16 62
2.5. Generalised trust and GDP per capita in European countries, 2002-14 68
2.6. Generalised trust and unemployment in European countries, 2002-14. 69
2.7. Generalised trust and life expectancy in European countries, 2002-14 69
2.8. Generalised trust and life satisfaction in European countries, 2002-14. 70
2.9. Institutional trust and perceptions of corruption in OECD countries, 2006-15 . . . 71
2.10. Institutional trust and GDP per capita in OECD countries, 2006-15 72
2.11. Institutional trust and unemployment rates in OECD countries, 2006-15 73
2.12. Institutional trust and homicide rates in OECD countries, 2006-15 74
2.13. Institutional trust and life satisfaction in OECD countries, 2006-15 75
4.1. The planning process: From user needs to survey questions 118
4.2. Generalised trust in the United States, 1972-2014 127
4.3. A system of well-being statistics: Conceptual overview 133
5.1. Trust in public service by gender in New Zealand. 159
5.2. Interpersonal trust in Sweden, 1996-2015 159
5.3. Interpersonal trust by age groups in New Zealand, 2014. 160
5.4. Interpersonal trust in European countries, 2013. 161
5.5. Trust in public institutions in European countries, 2013 162
5.6. Trust in government in the United States, 1958-2015. 163
5.7. Trust in national government in 2014 and changes since 2007, OECD countries . 163
5.8. Interpersonal trust by region of birth in Canada, 2013. 164

Follow OECD Publications on:

http://twitter.com/OECD_Pubs

http://www.facebook.com/OECDPublications

http://www.linkedin.com/groups/OECD-Publications-4645871

http://www.youtube.com/oecdilibrary

http://www.oecd.org/oecddirect/

This book has...

StatLinks

A service that delivers Excel® files from the printed page!

Look for the *StatLinks* at the bottom of the tables or graphs in this book. To download the matching Excel® spreadsheet, just type the link into your Internet browser, starting with the *http://dx.doi.org* prefix, or click on the link from the e-book edition.

© OECD 2017

Executive summary

The *OECD Guidelines on Measuring Trust* aim to assist data producers in collecting and reporting trust measures, and to support users of trust data in understanding different measurement approaches and their implications for analysis. They describe best practices in trust measurement, propose a core set of measures to form the basis for international comparisons, and encourage national statistical offices (NSOs) to include measures in their regular household surveys.

In particular, the Guidelines are intended to:

- Improve the international comparability of trust measures by providing guidance for NSOs and other data producers, grounded in best practice in question design;
- Summarise what is known about the validity and reliability of trust measures and, where possible, extend this evidence through empirical analysis of existing surveys;
- Act as a catalyst for NSOs and researchers to broaden the evidence base on the validity and reliability of trust measures;
- In the longer run, increase the number of countries for which official measures of trust are produced, so as to contribute to the monitoring of critical targets of the Sustainable Development Goals.

What is trust?

The Guidelines make a fundamental distinction between individuals' trust in other people (***interpersonal trust***) and trust in institutions (***institutional trust***):

- For interpersonal trust, the Guidelines distinguish between *generalised trust* and *limited trust*. Generalised trust refers to trust in people who are not known to the respondent or to trust in situations where the person being trusted is not specified. Limited trust focuses on persons known to the respondent, including family, friends and neighbours.
- Institutional trust refers to trust in all types of institutions, with trust in *political, law and order* and *non-governmental institutions* used to refer to a narrower concept.

Why have these Guidelines been produced?

The policy need for better measures of trust has been underscored by international initiatives such as the United Nations Sustainable Development Goals and work to improve measures of key economic, social and environmental outcomes, such as that initiated by the Report of the Commission on the Measurement of Economic Performance and Social Progress in 2009 and carried forward by the OECD's Better Life Initiative.

Trust measures based on household surveys are already collected as part of the official statistical system in several OECD countries. This has largely been the result of demand

from these countries' policy makers for better information on people's well-being, social capital and social cohesion. However, regular, timely and consistent measurement is less common, and most comparable information at the moment comes from unofficial sources.

How are the Guidelines intended to be used?

The Guidelines do not prescribe a single approach to measurement, but instead bring together information on what is currently known about measuring trust. Users of the Guidelines will find different parts of this document valuable depending on their needs.

- Chapter 2 addresses the issues of concept, relevance and validity. This chapter will be of interest to both users of trust data, interested in what concepts are captured by different measures and the extent to which different trust measures are valid, and to producers of trust data, wanting to reach a judgement about whether to measure trust and which concepts to focus on.
- Chapter 3 brings together information on the key methodological issues that should inform question design, including issues relating to measurement error, question wording, response formats, survey context, survey mode, response styles and cultural context. This chapter provides a resource for producers of trust data interested in question design as well as for technical users wanting to understand different sources of bias in trust data.
- Chapter 4 provides specific guidance on best practice in measuring trust. This chapter is structured around the different stages of the research process, i.e. planning, survey and sample design, question design, and implementation. More than the other chapters, this section is prescriptive, and is intended particularly for data producers during the survey design process.
- Chapter 5 focuses on the output and analysis of trust measures. The main goal for the chapter is supporting the production of basic descriptive outputs and the interpretation of trust data, such as evidence on what can be considered as *large* or *small* changes in trust levels. The chapter also reviews the analysis of trust data to help users who are approaching the data with a research question requiring more sophisticated analysis.

There are two annexes to the Guidelines.

- Annex A brings together a wide range of trust questions currently in use in different surveys around the world, from both official and unofficial surveys. This annex will help readers understand what sorts of measures are available from different sources.
- Annex B contains five prototype question modules as starting point for data producers to develop their own questions. The first core question module includes a limited set of measures that are intended for widespread use. A single 'primary measure' defines the absolute minimum that should be included in relevant surveys and forms the basis for cross-country comparison; another four questions provide basic information on the most important types of interpersonal and institutional trust. The other four modules cover different approaches to capturing information on trust, focusing on respondents' evaluation of their feeling of trust, expectations about others' behaviour, past experiences, and experimental techniques to observe trusting behaviour.

Next steps

As the evidence base on trust develops, the information in these Guidelines will eventually need to be updated. The Guidelines should thus not be viewed as "carved in stone". Ideally, a review should take place some time after their release to assess the degree to which the evidence base has improved and to identify any next steps to move towards greater standardisation of trust measurement in the official statistical system.

© OECD 2017

Overview and recommendations

The main points and recommendations from the Guidelines are summarised in the following. These are organised in four sections, each reflecting the content of the substantive chapters of the Guidelines. The section corresponding to Chapter 4 (measuring trust) is outlined in more detail as this chapter provides the most detailed and prescriptive recommendations relating to the collection of trust data.

1. Concept and validity

The relevance of measures of trust is not in doubt. Measures of interpersonal trust – particularly generalised trust – are of fundamental importance to assessing the well-being of societies, to measuring social capital, and to understanding the drivers of other social and economic outcomes. This is reflected both in the large and growing literature on the drivers and consequences of generalised trust and in the wide range of national and international initiatives that include generalised trust as an outcome of interest, ranging from national initiatives to measure well-being such as those of the United Kingdom and Israel through to the UN Sustainable Development Goals. Similarly, there is a high level of interest in institutional trust, which is reflected both in measurement initiatives and in a large and varied academic literature. In addition to being important for measuring well-being, understanding institutional trust is essential to understanding government effectiveness and the functioning of democratic systems of government.

The accuracy of trust measures is less clear. In general, while evidence on the validity of measures of generalised trust is strong, there is relatively little evidence focusing on the validity of limited trust measures. Measures of generalised trust perform well in terms of face validity, construct validity and convergent validity. This holds whether the measures are assessed at the cross-country level or at the level of individual responses. While there is some question over the test-retest reliability of some measures of generalised trust at the individual level, country-level results are highly reliable across different data sources and over time.

As with all intangible concepts, measuring generalised trust raises a number of issues about respondents' interpretation of the question in front of them and about their subjective judgement, but these issues are not intractable. There is good evidence that, despite these issues, questions on generalised trust produce valid data, and extensive research is providing new insights into the remaining measurement issues. In fact, it is likely that more is understood about the validity of measures of generalised trust than is the case for many more traditional elements of official statistics. From this perspective, measures of generalised trust can be considered as fit for purpose and should be measured in official surveys where relevant.

The picture for measures of institutional trust is more mixed than is the case for measures of interpersonal trust. While institutional trust measures generally perform well in terms of construct validity, the situation is less clear with respect to face validity and convergent validity. There is thus reason to believe that such measures might be biased in some circumstances, and for a number of key aspects of validity there is simply no evidence one way or the other. For this reason, despite the policy relevance of measures of institutional trust, it is less easy to state that they are fit for purpose within the context of official statistics. This does not mean that such measures have no place in the official statistical system, however. Rather, it suggests that such measures should be considered more experimental and should be implemented in contexts where their experimental status is clear; this is particularly important for national statistical offices.

On the other hand, the relevance of such measures suggests that they should be a high priority for further research, both in the academic community and within national statistical agencies. Many of the key methodological questions regarding the accuracy of measures of institutional trust will require the sample size and response rates that only national statistical offices are able to deliver.

2. Methodological considerations in the measurement of trust

This section examines the impact of different sources of bias and measurement error on trust questions. While trust measures are more sensitive to response biases than more objective measures (such as educational attainment or life expectancy), these biases are also likely to occur in other self-reported measures that are already being collected by NSOs. Although it is essential to be aware of these biases and of the most appropriate question and survey design strategies to mitigate them, the existence of measurement error *per se* is not an argument against gathering data on trust. No matter which approach to question design is adopted by data collectors, standardisation is critical to ensure meaningful comparison over time and between groups and countries.

Question wording

The evidence on question wording (especially that drawn from split sample experiments) shows that good question wording matters for results.

- Question wording should avoid referring to concepts other than trust and be specific and precise to the situation of interest.
- For interpersonal trust, a neutral question wording is recommended: data collectors should refrain from referring to "caution in dealing with other people", as this wording can prime more vulnerable groups to report lower trust.
- For institutional trust, specifying what institutions are expected to do can make a difference to respondents in some cases.
- Overall, question wording should be precise enough to be understood by respondents, without getting into subtle nuances that might also pose problems for translatability across countries. If the concepts that different questions try to capture are too narrowly related, respondents might have difficulty differentiating between them (e.g. trust vs. confidence).

Response formats

The way response options are presented can have a significant impact on the distribution of responses.

- For trust items, a numerical 0-10 scale with verbal scale anchors is recommended, as it allows for a high degree of variance in responses, increases overall data quality and facilitates translatability across languages.
- The response order should be presented consistently (i.e. 0-10 instead of 10-0) in order to minimise mental switching between positive and negative normative outcomes.
- Verbal descriptions of the scale anchors should represent absolute responses (e.g. *completely/ not at all*) to minimise acquiescence bias and socially desirable responding and to allow for the full spectrum of possible responses.

Survey context

Trust measures should be considered within the broader survey context in which they are placed. As with the standardisation of wording and response formats, consistency of order within question modules across surveys and over time is essential to guarantee the quality and comparability of trust measures.

- Since order effects occur most often when two or more questions deal with the same or closely related issues, trust items should either be separated within the survey as much as possible or buffered by intervening text.
- Whenever lists of trust items are used, a survey should move from a broad to a narrow level of specificity within a group of questions, e.g. by placing items about generalised trust before questions about limited trust.
- Generally, trust questions should not be asked immediately after items that are likely to elicit strong emotional responses or that refer to experiences with other people or institutions.
- Questionnaire designers should equally reflect on the potential effect that trust questions themselves can have on subsequent items, in particular those dealing with similar content.
- To minimise the impact of holidays, seasons and elections, data collection is recommended to be spread throughout the year or at least over multiple weeks.

Survey mode

Evidence suggests that trust questions can be sensitive, triggering respondents to answer in socially desirable way or be unwilling to answer at all. This is especially true for measures of trust in institutions.

- Self-administered surveys, compared to interviewer-led ones, perform better in terms of minimising social desirability.
- In all survey modes, sensitivity-related response biases can be reduced by decreasing the respondent's concerns about data protection (e.g. via confidentiality assurances) or by controlling the survey situation (e.g. not having enumerators give out information about their own social identity).
- If face-to-face interviews are the only option, the use of innovative interviewing methods such as the sealed envelope or unmatched count technique could be explored.

Response styles and cultural context

Cross-cultural response styles are very difficult to verify externally against a common standard or actual behaviour. Even where the existence of response styles has been established, they do not necessarily harm overall data quality.

- If data producers want to mitigate the possibility of response style bias, they should, rather than relying on *ex post* statistical adjustment techniques, focus on designing the questionnaire so that items are as simple, easy to interpret and minimally burdensome as possible.
- The overall survey design (including its length and how it is introduced) needs to pay particular attention to respondent burden, motivation and fatigue in order to maximise data quality.

- Question formats that are more prone to response biases should be avoided: for example, *agree/disagree* and to a lesser degree *yes/no* response formats are more likely to prompt acquiescence.

Issues for further research

Further research is needed on both institutional and interpersonal trust, but especially on the former, for which there is very little methodological evidence available.

- First, with regard to question wording for institutional trust, experimental testing should be used to establish which specifications (e.g. *to act in the national interest*, or *to improve the life of someone like me*, or *to do what is right*) matter the most for which institutions. Ideally, these experiments should be carried out across more than just one country.
- Second, while it has been suggested that the use of a mixture of positive and negative statements can mitigate both *yay* and *nay* saying, this approach needs to be further tested to rule out the risk of confusing respondents when the same scale end presents something positive in one item and something negative in a following one.
- Third, with regard to order effects, it is not yet clear in which cases these occur for trust questions. More targeted methodological research is needed to discover what type of question or what type of context triggers which effect in order to further inform survey design. While there is some evidence that transitional text between questions can act as a buffer to mitigate order effects, various text versions should be tested for their impact on trust questions.
- Fourth, more research that validates response styles from different cultures against external references, such as actual trusting behaviour in real life or experimental games, would enrich the current body of cross-cultural trust research.

3. Measuring trust

The aim of this section is to outline best practice in the measurement of trust. This covers decisions about what to measure and the best approaches to measurement. Issues of sample design, survey design, data processing and questionnaire design are covered here.

Planning for the measurement of trust

- Decisions about what to measure should always be grounded in a clear understanding of user needs. Important questions to consider include: i) What are the policy questions?; ii) Is the trust content being proposed appropriate to respond to the policy questions?; iii) Does the measure proposed allow monitoring changes over time or comparing population groups?; iv) What population groups are of greatest interest to the user?; v) Does the user's interest lie in comparing outcomes of different groups or in understanding the relationship between different aspects of trust?; vi) Is the user's primary interest in generalised trust, limited trust or institutional trust? If the focus is on the latter, which institutions are of primary interest?; vii) What are the frequency requirements of the users to monitor changes over time?; and viii) What within-country comparisons are required, such as geographic level?
- It is imperative to consider not only how best to measure trust *per se*, but also what other measures should be collected alongside measures of trust for analytical purposes. These should include: i) Age; ii) Gender; iii) Marital status; iv) Household type; v) Presence of

children; vi) Household size; vii) Geographic information; and viii) Migration status/ Country of birth/Year of arrival.

- In addition to the demographic measures identified above, which can be considered essential, a number of additional variables may also be useful: i) Language spoken at home; ii) Living in urban/rural areas; iii) Income; iv) Wealth; v) Employment status; vi) Educational attainment; vii) Health status; viii) Social contact and networks; ix) Civic engagement and governance; x) Personal security and victimisation; xi) Subjective well-being; xii) Ethnic identification; and xiii) Religion.

Survey and sample design

Sampling

- Responses to questions on trust are inherently personal, and consequently the unit of measure must be the individual. This implies that the sampling frame must produce a representative sample of individuals or households as if all individuals are personally interviewed.
- In general, measures of trust would be collected for the entire adult population (aged 15 and older).

Frequency of data collection and duration of enumeration

- It is not possible to provide specific guidelines for how frequently measures of trust should be collected that cover every contingency, since the range of possible data uses is large and the frequency at which data are needed will vary depending on the intended use and on the type of measure in question.
- For the purposes of monitoring well-being and for assessing trends in social capital, an annual time series should be regarded as the minimum in terms of frequency of enumeration.

Duration of enumeration

- The duration of the enumeration period (i.e. the period of time over which information is collected) is important for measures of trust. Unlike measures of educational attainment or marital status, for which it does not usually matter at what point during the year the data are collected, the precise timing of the collection period might have an impact on measured trust.
- Ideally, enumeration of trust data would take place over a full year and would include all days of the week, including holidays. This would ensure that measures of trust provide an accurate picture for the whole year. Where a year-long enumeration period is not possible, enumeration should, as far as is possible, be spread proportionately over all days of the week.

Sample size

- Large samples are highly desirable in any survey: they reduce the standard error of estimates and allow both a more precise measure of trust as well as a greater degree of freedom when producing cross-tabulations and analysis of results for population sub-groups. With measures of trust, sample size is particularly important because of the relatively small changes in trust associated with many areas of analytical interest.

Survey mode

- In terms of data quality, Computer-Assisted Self-Interviewing/Computer-Assisted Personal Interviewing (CASI/CAPI) with show cards should be considered best practice for collecting trust data. The presence of an interviewer allows for a strong rapport to be built with the respondent, while show cards help with data quality.
- The confidentiality provided by CASI sections to the interviewing should help address respondent reluctance to provide accurate answers to potentially sensitive questions.
- Where other modes are used, it is important that data producers collect information to enable the impact of mode effects to be estimated. National statistical agencies, in particular, should consider experimentally testing the impact of the survey mode on responses to the core measures of trust and publish the results along with any results from CATI or CASI surveys.

Survey vehicle

- Where trust, governance or social capital are the key area of interest, it may be appropriate to build a special module focused specifically on trust. This is especially the case where the use of trust data focuses on measuring social capital or on evaluating governance.
- As trust measures are of analytical interest in a broad range of different contexts, a limited range of trust questions can usefully be included in a wide range of surveys.

Question placement

- Important trust questions should be included in the core section of the survey. Although it is not possible to place trust questions at the start of every survey, the effect of bias due to context effects can be limited if trust questions are included in a fixed portion of the survey questionnaire. While this does not eliminate bias, it will not affect analysis of differences in levels across population groups or over time.
- Trust questions should not be placed immediately after questions that are likely to prime respondents with regard to trust, or that respondents might use as a heuristic for determining their response to the trust question. This includes questions on social contact, victimisation, political beliefs, risk or insecurity. The best questions to precede trust questions are demographic questions.
- Transition questions should be used to refocus respondent attention. However, it is important to consider the risk that transition questions might introduce their own context effects. For example, drawing attention to a respondent's personal life may lead them to focus on personal relationships rather than on strangers when answering subsequent questions about interpersonal trust. Development of effective transition questions is a priority for future work.
- Introductory text should be used to distinguish between question topics. Well-worded text preceding each question or topic can serve as a buffer between measures of trust and sensitive questions. Further cognitive testing or experimental analysis of the impact of different types of introductory text would be of high value.

Question order

- In terms of ordering question modules, evidence suggests that moving from the general to the specific is the best approach. For interpersonal trust, a question on generalised trust should be placed ahead of more specific questions relating to limited trust. Also,

questions on interpersonal trust should be asked before questions on institutional trust, which tend to be more specific and are also likely to have a significant priming effect.

- Questions on trust in institutions should proceed from better-known institutions to more obscure ones. Where comparisons between levels of trust in different institutions are important, the order in which questions are presented should be randomised for each respondent. If this is not possible for the whole sample, pilot testing should involve the randomisation of question order so that the size of any bias in measured trust is known.

Translation

- Initial translation should ideally be carried out by at least two independent translators who have the destination language as their mother tongue and who are fluent in the source language. Translators should be informed about the goal of the study and be familiar with the background, origin and technical details of the source questionnaire, as well as with the nature of the target population. As with any survey design, cognitive interviewing and field testing should be undertaken, with the results reviewed before the full survey goes into the field.

Questionnaire design

- These Guidelines provide five prototype question modules for the measurement of trust. Module A contains a set of ***core measures***, which include a single primary measure of generalised interpersonal trust that is intended to form the baseline for international comparisons, and is the highest priority for inclusion in any attempt to measure trust.
- Module A contains those questions for which evidence of validity and policy relevance is the strongest and which are most apt to achieve some degree of international harmonisation. Unlike all of the other question modules included in these Guidelines, the core module is intended to be used without significant amendment and in full.
- Modules B to E are focused on different approaches to measuring trust. These modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires.

Survey implementation

- Interviewer training is crucial to the quality of responses in any survey. To manage risks around respondent attitudes to questions on trust, interviewers should be well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used.
- Evidence suggests that measures of trust are relatively non-problematic for respondents to answer. Item-specific non-response rates for interpersonal measures of trust are similar to those for marital status, education and labour market status, and much lower than for those on household income. Non-response rates for questions on institutional trust are somewhat higher, but still lower than is the case of income. This suggests that, in general, trust questions are not perceived as problematic by respondents.
- Normal data-cleaning procedures include looking for obvious errors such as transposed numbers, duplicate records, loss of records, incomplete responses, out-of-range responses or failure to follow correct skip patterns. Some issues are of particular relevance to trust data. In particular, where a module comprising several questions with the same scale is used, data cleaning should also involve checking for response sets (see Chapter 3).

4. Output and analysis of trust data

The output and analysis of measures of trust is inherently complex and difficult to reduce to a succinct series of recommendations. This section draws together general advice on the output of trust data along with some information on interpreting results and on issues involved in undertaking analysis of trust data.

Reporting trust data

Analysts tasked with reporting trust data have an important communications role to play and should take into account the intended target audience and its needs. There are several ways to output trust data, each with unique pros and cons.

- Trust levels can be presented by reporting frequencies in each category, the proportions above or below a given threshold, or central tendency measures (i.e. *mean*, *median*, *mode*).
- Some rules of thumb for best reporting practice include refraining from arbitrary labels for thresholds (e.g. *high*, *low*) and complementing mean levels with information about the distribution of data, such as the standard deviation.
- Changes over time can be monitored by tracking changes in mean trust through time series or by calculating changes in the mean score over various points in time.
- Group differences can be examined by presenting group differences over time, relative to a given threshold, or by showing the (absolute or percentage) differences in the proportion of respondents who have selected a specific answer.
- Both sample size and standard errors should be reported alongside group means.

Interpreting trust data

Essential questions for the interpretation of trust data deal with what should be considered a *small* or a *big* difference between observations in real-life terms, and the extent to which observed differences are influenced by measurement artefacts and errors. While these Guidelines provide an initial attempt to document the magnitude of differences (between population groups, between countries, over time) encountered up to now with existing data, many gaps remain, and knowledge on the overall data universe of trust will remain limited until higher quality and more frequent data become available. The magnitude of differences between and within observations (over time), can be influenced by a variety of factors that should be factored into any interpretation exercise. These include:

- the limits imposed by the response scale
- issues of reverse causality
- the possible impact of culture on trust estimates.

Analysis of trust data

Sound analysis of trust data requires access to data from which causal inferences can be made and that the relevant covariates, including standard demographic and control variables, are ideally collected in the same survey. The choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – matters. In addition, community characteristics often influence trust levels, and aggregate trust levels influence individual-level well-being outcomes.

Keeping common econometric challenges in mind when working with trust data is very important. These include omitted variable bias, over-identification, reverse causality and shared method variance.

5. Next steps

These Guidelines do not aim at providing the final word on the measurement of trust or at developing a formal international standard for measuring trust. While formal statistical standards are an important part of official statistics, it is appropriate to produce them only when the measure in question is well understood and when there is a well-developed body of statistical activities pursued by NSOs, conditions that are not currently met in the case of trust. These Guidelines aim, rather, to bridge the gap between, on the one hand, the current situation of scattered measurement in official statistics and the more widespread – but still inconsistent measurement elsewhere – and a formal standard, on the other. By encouraging NSOs to collect trust data more systematically, i.e. by including a core set of internationally comparable questions in their surveys, the Guidelines aim to support the development of the evidence base that might in the future underpin decisions about developing an international statistical standard for measuring trust.

© OECD 2017

Chapter 1

Introduction

This chapter describes the context in which these Guidelines were produced. The first part of the chapter sets out the motivation for producing these Guidelines and identifies some of the key international initiatives that are creating a need for better trust measures. The second part presents the scope and objectives of the Guidelines and provides an overview of the structure and contents of the full report.

1.1. Introduction

Trust is a concept of fundamental importance to the well-being of individuals, and to society more broadly. At the individual level, in order to live comfortably, people need a personal feeling that the other members of the community with whom they interact daily can be trusted. At a societal level, trust is essential to the smooth functioning of society: every day-to-day transaction that we make involves some degree of trust in the people with whom we interact, and it is the trust in these interactions that supports the prosperity of the world's economy. As argued by Nobel Laureate Kenneth Arrow: "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence" (Arrow, 1972, p. 357).

It is not just trust in other people that matters. Trust in institutions also underpins a successful society. Without some degree of trust in institutions such as the parliament, the civil service, the justice system and the police, it would be impossible for a community to perform effectively or for the individuals within a country to live the sort of lives that they wish to pursue. Trust in institutions requires that these institutions are competent and effective in delivering on their goals, but also that they operate consistently with a set of values that reflect citizens' expectations of integrity and fairness.

Although widely acknowledged as important, the measurement of trust does not have a long tradition, particularly within official statistics. This partly reflects a paucity of evidence on the validity and reliability of different measures of trust, as well as – until recently – a lack of strong policy demand for such metrics. However, this situation has begun to change radically. A range of recent policy initiatives has underscored the urgent need for better measures of trust – both in other people and in institutions – to help inform policies aimed both at improving the quality of governance, social cohesion and citizen engagement and at better understanding the drivers of well-being and economic performance.

The OECD Guidelines on Measuring Trust aim to make the concept of trust measurable through official statistics by describing best practices in its measurement, by proposing a core set of measures that could form the basis for international comparisons, and by encouraging national statistical offices (NSOs) to include measures of trust in some of their regular household surveys. The Guidelines are intended to assist data producers in collecting and reporting measures of trust as well as to support users of trust data in understanding the different approaches to measurement and the implications of these for analysis.

1.2. Motivation

Recent initiatives

Concern about levels of trust in society is highly topical. In the aftermath of the 2008 financial crisis, people's trust in a broad range of public and private institutions (e.g. banks) fell sharply in most OECD countries, especially those most severely affected by the crisis. In this context, OECD ministers, at the 2013 OECD Council meeting on Jobs, Equality and Trust,

called upon the OECD to strengthen its efforts to better understand how trust in public institutions shapes economic performance and people's well-being, as well as to identify its key drivers. This mandate led to the OECD Trust Strategy, a two-year initiative aimed at providing methodological, empirical and practical guidance to OECD governments wanting to restore people's trust in public institutions.

The importance of developing reliable measures of trust has been independently underscored as part of the OECD Better Life Initiative. Since 2011, measures of trust in others have been used as a measure of current social capital in the biannual *How's Life?* report, while trust in institutions has been considered in the same report as an indicator of civic engagement and governance (one of the 11 dimensions of current well-being identified in the initiative). Since the 2015 edition of *How's Life?,* both interpersonal and institutional trust feature as measure of social capital (in the chapter on "Resources for future well-being"). When discussing the statistical agenda ahead in the field of "social connections", the OECD (2011) stated that "it is important to ensure the standardisation and regular collection of official statistics for those measures for which there is a sufficient body of knowledge, in particular with respect to interpersonal trust". Similarly, the OECD (2011) identified a number of actions aimed at improving measurement in the area of civic engagement and governance and argued that:

> …better measures are required to assess how people perceive the quality of the democratic institutions in the country where they live. While many unofficial surveys contain questions for measuring civic engagement and governance they typically have a narrow geographic coverage, small sample size and inadequate sampling procedures. Steps should be taken to include questions on civic participation and trust in institutions in large-scale official surveys, through a combination of a few recurrent questions in regular surveys and more detailed questions in (less frequent) dedicated modules, as is already done in some OECD countries (e.g. special modules of the Current Population Survey on voting and civic engagement in the United States).

Looking beyond the OECD, the statistical community today confronts a clear imperative to improve the measurement basis of the UN Sustainable Development Goals (SDGs) agreed in September 2015 by the UN General Assembly. In particular, Goal 16 of the SDGs ("Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels") focuses explicitly on governance, institutional quality and political participation. Even when not explicitly mentioned, trust in others and in institutions is an essential prerequisite for the types of behaviours and policies that are needed to preserve those public goods and global commons that underpin inclusive and sustainable growth.

In response to this political demand, the statistical community has launched a range of new measurement initiatives (United Nations, 2015). In September 2014, representatives of NSOs and experts convened in Cape Verde to develop draft terms of reference for a City Group on Governance Statistics to be established under the auspices of the UN Statistical Commission. The creation of such a group was agreed by the UN Statistical Commission in March 2015, and this group – the Praia Group – is today developing a handbook of governance statistics as well as advising on governance measures related to Goal 16 of the SDGs. The programme of the Praia Group involves multiple work-streams covering different elements of governance, including people's trust in public institutions and their experiences with the performance of various institutions and with corruption. Given the scope of the Praia Group's

mandate, the group partly functions as a clearing house, bringing together advice and guidance from different sources on measuring various aspects of governance. In this context, these Guidelines on Measuring Trust are meant to constitute one of the OECD's primary contributions to the Praia Group.[1]

Beyond the contribution of better measures of trust to the specific initiatives outlined above, these Guidelines are part of a broader OECD work programme aimed at improving the measures of key economic, social and environmental outcomes. Following the release of the *Report by the Commission on the Measurement of Economic Performance and Social Progress* in 2009 (led by Amartya Sen, Joseph Stiglitz and Jean-Paul Fitoussi), the OECD has played an important role in carrying forward the Commission's measurement agenda. This has included applying the Commission's recommendations to regularly monitor and analyse well-being and sustainability across the OECD area through the *How's Life?* series of reports (OECD, 2011, 2013a, 2015 and 2017a) and to engage ordinary people through specific communication tools (the Better Life Index). It also includes pushing forward the frontiers of measurement by producing guidelines for measuring subjective well-being, the distribution of household wealth and the quality of the working environment (OECD, 2013b, 2013c, 2017b). It is in this context that these Guidelines on Measuring Trust should be seen: as part of a broader effort aimed at building a coherent system of statistics on well-being.

The need for guidelines

For official statistics, statistical quality is grounded in the commitment of NSOs to follow best practice in data production. This involves the use of consistent international concepts, classifications and methods. The *UN Fundamental Principles of Official Statistics* (2014) together with the enabling statistical legislation of many countries enshrine this goal in written form. However, official statistical standards do not simply emerge fully formed: they can be produced only when there is a sufficient body of understanding around the measures in question to support clear judgements about best practice. Thus, the development of formal standards necessarily begins in experimental and informal data collection.

Trust measures based on household surveys are already collected as part of the official statistical system in several OECD countries. This has largely been the result of demand from policy makers for better information on well-being, social capital and social cohesion. However, regular, timely and consistent measurement is much less common. In most cases, such as for many European Union countries, official statistics on trust have been collected only through an ad hoc module to one of the main Community surveys (the 2013 EU-SILC well-being module). Nevertheless, some European countries have started to implement national well-being modules of EU-SILC and recurrent general social surveys that include trust measures (such as Poland's Social Cohesion Survey or the Social Cohesion and Well-being Survey of the Netherlands). In Australia, Canada and New Zealand, trust measures have also been collected by NSOs as part of a regular general social survey. The Mexican NSO INEGI has been looking at trust through its National Survey of Victimization and Perception of Public Security.

Although some official data on trust are already available, most comparable information comes from unofficial sources. The most significant of these are the Gallup World Poll (for trust in institutions only), the World Values Survey, the European Social Survey, the European Quality of Life Survey and the Barometers carried out in many countries around the world. Questions on interpersonal trust are also included in the OECD Survey of Adult Skills (PIAAC) and have been used to explore how skill acquisition helps to maintain and

increase people's trust in others. These surveys have allowed significant progress in understanding the degree to which trust measures are capable of providing useful information (discussed in Chapter 2). However, the available data still face significant limitations. In particular:

- Coverage from existing unofficial surveys is uneven, both across countries and over time.
- Most unofficial data are based on small samples (typically, of around 1 000 per country), which precludes looking at population sub-groups within each country.
- Available official data are not comparable across countries and are often collected on an ad hoc basis.
- Basic methodological information is lacking on which measures are most useful for different purposes, limiting the degree to which measures of trust can be used to inform policy.

These Guidelines on Measuring Trust aim to contribute to addressing these issues by setting out best practice, providing direction for methodological research, and encouraging NSOs both to collect more trust data and to do so in an internationally comparable manner. It is with these goals in mind that the OECD devoted resources to its production. It is hoped that these Guidelines will, in turn, contribute to a step change in the quality and availability of data on trust.

1.3. The Guidelines for measuring trust

Scope and objectives

These Guidelines are aimed, in the first instance, at NSOs that are either already measuring trust or that are contemplating new trust measures through additions to existing surveys or new data collection. In the former case, the Guidelines can provide guidance on which measures could be used to produce internationally comparable data and on how to present these data, while in the latter case much of the information on validity and methodological issues will also prove of relevance. NSOs are the primary audience for these Guidelines, both because the OECD is an inter-governmental institution where NSOs are directly participating in the work of the Organisation through the Committee on Statistics and Statistical Policy, but also because policy-relevant trust data will require the large sample sizes and high levels of statistical quality that only official data can provide.

Beyond NSOs, other producers of trust data will also find these Guidelines of value. As noted above, much of the trust data currently available come from non-official sources. Better and more comparable trust data from non-official sources will remain the primary source of information for some years to come, until a significant body of official data begins to accumulate. Also, much academic work will continue to use non-official data because they offer both greater flexibility with respect to adding in new variables and easier access to microdata. Finally, these Guidelines will also be of value to data users, not just by providing information on the use and analysis of trust data but also for the evidence on the validity of different trust measures and on the impact of different measurement methodologies on reported trust.

The scope of these Guidelines is limited to measures of trust by individuals in other individuals and in institutions. This is partly driven by practical concerns – the Guidelines are aimed at NSOs and the sort of data that they can collect – but also because it is the trust of individual citizens that is of greatest policy interest. Hence, the main focus of the

Guidelines is on measures of generalised interpersonal trust, which captures the trust of individuals in people who are not known to them, along with trust in public institutions such as the police, parliament and the justice system. A more detailed discussion of the concept of trust and the specific aspects of trust that can be measured is the focus of the first part of Chapter 2.

It should also be noted that the scope of these Guidelines goes beyond providing advice on best practice for measuring trust. A strong focus is on assessing the relevance and accuracy of trust measures, particularly with respect to validity. This is not just because validity is of importance to potential data producers in assessing whether it is worthwhile to collect a measure, but also because there are significant knowledge gaps in this area. The Guidelines aim both to fill some of these gaps where possible and to indicate strategic priorities with respect to trust research for the academic research community.

A key element of these Guidelines is a set of prototype question modules on trust that can be used as a starting point by data producers in developing their own measures, and which will provide a basis for international comparison. In particular, the Guidelines are intended to:

- Improve the international comparability of trust measures by establishing common standards that could be used by NSOs and other data producers, grounded in best practice in question design.
- Summarise what is known about the validity and reliability of measures of trust and, where possible, extend this body of information through empirical analysis of existing survey and proxy measures of trust in government.
- Act as a catalyst for broader work by NSOs and researchers to broaden the evidence base on the validity and reliability of trust measures.
- In the longer run, increase the number of countries for which official measures of trust are produced, so as to contribute to the monitoring of critical elements of the SDGs.

As a last point, these Guidelines do not aim at providing the final word on the measurement of trust or at developing a formal international standard for measuring trust. While formal statistical standards are an important part of official statistics, it is appropriate to produce these only when the measure in question is well understood and when there is a well-developed body of statistical activities pursued by NSOs, conditions that are not currently met in the case of trust. The Guidelines aim, rather, to bridge the gap between, on the one hand, the current situation of scattered measurement in official statistics and the more widespread, but still inconsistent measurement elsewhere, and a formal standard on the other. By encouraging NSOs to collect trust data more systematically, including a core set of internationally comparable questions in their surveys, the Guidelines aim to support the development of the evidence base that might in the future underpin decisions about developing a true international statistical standard for measuring trust.

As the evidence base on trust develops, the information in these Guidelines will eventually become outdated. For this reason, it is not envisaged that the Guidelines should be viewed as "carved in stone". Ideally, a review of these Guidelines should take place some time after their release to assess the degree to which the evidence base has improved subsequent to publication. This review should assess whether the Guidelines themselves are in need of revision and identify any next steps that should be taken in order to move towards greater standardisation of trust measurement in the official statistical system.

The structure of the Guidelines

These Guidelines include four substantive chapters and two annexes. Chapter 2 addresses the issues of concept, relevance and validity. For meaningful measurement to be possible, it is essential to have a clear understanding of what is to be measured. The first part of Chapter 2 addresses this topic by reviewing the main definitions of trust found in the academic literature and discussing how these relate to each other. A working definition of trust that covers both interpersonal and institutional trust is proposed; it forms the basis for the survey questions proposed by these Guidelines for measuring trust. A brief account of the main approaches to measuring trust is also provided, which is used as the basis for organising discussion around measurement later in the Guidelines. The second half of Chapter 2 focuses on statistical quality, with a particular emphasis on the issues of relevance and validity. This section aims to provide a clear account of why and how measures of trust are useful to policy makers and the general public and sets out what is known about the accuracy of trust measures.

Chapter 3 brings together information on the key methodological issues that should inform the design of questions on trust. The chapter is structured under broad headings relating to measurement error, question wording, response formats, survey context, survey mode, response styles and cultural context. Each of these issues is wide-ranging, and the Guidelines do not attempt to cover all aspects of these issues in depth. The main focus in each case is on summarising findings that apply specifically to measures of trust. Beyond this, where there are issues that are likely to be particularly important to designing trust questions but for which no trust-specific literature exists, a brief summary is provided, drawing evidence from other areas (such as the design of subjective questions more generally). As users of the Guidelines are assumed to have a good grounding in the essentials of question design, the aim here is to summarise information that is particularly relevant to designing questions on trust, rather than starting from scratch.

Chapter 4 provides specific guidance on best practice in measuring trust. This chapter is structured around the different stages of the research process, covering planning, survey and sample design, question design, and implementation. More than the other chapters in these Guidelines, this chapter is prescriptive. In particular, it provides specific recommendations on issues such as sample size, frequency and duration of enumeration and provides the rationale behind the specific choice of questions included in the prototype question modules attached to the Guidelines.

Chapter 5 focuses on the output and analysis of trust measures. The main goal for the chapter is supporting basic descriptive outputs of trust data, including examples of how such data have been reported in the past. The chapter also covers the interpretation of trust data, such as evidence on what can be considered as *large* or *small* changes in trust levels. The chapter also reviews the analysis of trust data to help users who are approaching the data with a research question requiring more sophisticated analysis.

There are two annexes to the Guidelines. Annex A brings together a wide range of trust questions currently in use in different surveys around the world, sourced from both official and unofficial surveys. This annex is intended to serve as a reference point for the discussion of questions from specific surveys in the main text of the Guidelines and as a resource for analysts wanting to know about the range of available measures. It should also help in gaining an understanding of what sorts of measures are available from different sources. Annex B also features questions on trust but has quite a different focus. It contains five

prototype question modules on trust that are intended to be used by data producers as a starting point for developing their own questions. The first question module – the core module – includes a limited set of measures that are intended for widespread use. A single "primary measure" defines the absolute minimum that should be included in relevant surveys and forms the basis for cross-country comparison; another four questions – also intended to be used as written – round out the module and provide basic information on the most important types of trust. The other four modules each cover a different approach to capturing information on trust. Unlike the core module, these modules are not intended to be used as written, but rather as a resource for data producers in designing their own questionnaires.

1.4. Conclusion

Key points made in this chapter are as follows:

- The OECD Guidelines on the Measurement of Trust aim to make trust measurable by setting out best practice in the measurement of trust, providing guidance towards a core set of measures that can form the basis for international comparisons and encouraging NSOs to include measures of trust in some of their regular household surveys.
- There is currently a strong policy need for better measures of trust, underscored by international measurement initiatives such as the United Nations Sustainable Development Goals and work aimed at improving the measures of key economic, social and environmental outcomes, such as that initiated by the *Report of the Commission on the Measurement of Economic Performance and Social Progress* in 2009 and carried forward by the OECD through its Better Life Initiative.
- The Guidelines are aimed at producers and users of trust data, with a particular emphasis on NSOs. The main focus is trust by individuals in other individuals and measures of trust in public institutions.
- These Guidelines are not a formal international standard, but are aimed at supporting data producers in their own initiatives to measure trust.
- As high-quality official measures of trust become available, it will be necessary to review these Guidelines in light of the new information. It is envisaged that a review of the Guidelines will take place once a sufficient body of national experiences has emerged, with recommendations for next steps in the measurement of trust.

Notes

1. OECD contributions to the Praia Group also include the statistical work undertaken in the context of the bi-annual report *Government at a Glance*, inputs provided to the in-depth review of government statistics carried out by the Conference of European Statisticians in 2016, and the special chapter on "Governance and Well-being" in the 2017 issue of *How's Life?*.

References

Arrow, K.J. (1972), "Gifts and exchanges", *Philosophy and Public Affairs*, Vol. 1, No. 4, John Wiley & Sons, Inc., pp. 343-362.

OECD (2017a), *How's Life? 2017: Measuring Well-being*, OECD Publishing, Paris.

OECD (2017b), *OECD Guidelines on Measuring the Quality of the Working Environment*, OECD Publishing, Paris.

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

OECD (2013a), *How's Life? 2013: Measuring Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264201392-en*.

OECD (2013b), *OECD Guidelines for Micro Statistics on Household Wealth*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264194878-en*.

OECD (2013c), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

OECD (2011), *How's Life? Measuring Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264121164-en*.

Stiglitz, J.E., A. Sen and J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, *www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf*.

United Nations (2015), "Transforming our world: The 2030 Agenda for Sustainable Development", *A/RES/70/1*, United Nations General Assembly, New York.

United Nations (2014), *Fundamental Principles of Official Statistics*, A/RES/68/261, United Nations, Geneva and New York.

© OECD 2017

Chapter 2

Concept and validity

A clear definition of the concept of trust is necessary in order to measure it. Building on a review of different theoretical approaches to the concept of trust, this chapter provides a working definition and a conceptual framework to underpin its measurement. The framework distinguishes between interpersonal and institutional trust and between the two main categories of interpersonal trust: trust in others, and trust in friends, family and neighbours. Four approaches to measuring trust are identified, which provide the organising framework for the question modules proposed in these Guidelines. The chapter also discusses the statistical quality of trust measures, with a focus on their relevance and accuracy. While measures of both interpersonal and institutional trust are of high relevance, the picture is mixed with respect to accuracy. In particular, while there is strong evidence supporting the reliability and validity of measures of trust in others, the evidence base is weaker with respect to trust in friends, family and neighbours, and, even more so, with respect to trust in institutions. For this reason, the chapter recommends that measures of trust in others should be included in official statistics, while measures of trust in friends, family and neighbours and of trust in institutions should be regarded as more experimental.

2.1. Introduction

This chapter focuses on the definition of trust and summarises the available evidence relating to the statistical quality of measures of trust. The first part of the chapter outlines a conceptual framework for measuring trust: before trust can be measured, it is essential to have a clear understanding of the concept(s) to be measured and of how this relates to different measurement approaches. Following this, the second part of the chapter considers issues of statistical quality, with a particular focus on the relevance and accuracy of trust measures. Relevance and accuracy are the focal aspects of statistical quality considered here because they are the elements of the *Quality Framework and Guidelines for OECD Statistical Activities* (OECD, 2008) that are most relevant to the issues covered by these Guidelines.

The conceptual framework developed here addresses the nature of trust and provides a model of the different types of trust that could be measured. This provides a framework for thinking about the most relevant measurement approaches and for identifying those types of trust that will be the focus for these Guidelines. Of the different types of trust identified in the conceptual framework, interpersonal trust and institutional trust are identified as falling within the scope of the Guidelines. Within these broad categories, trust in others (i.e. generalised trust) is identified as the most important element of interpersonal trust. A typology of approaches to measuring trust is also developed; this forms the basis for the different question modules proposed by the Guidelines, as described in Annex 2.

The largest part of this chapter focuses on issues of statistical quality, particularly on relevance and accuracy. Relevance concerns the degree to which measures of trust are sufficiently useful to justify collecting them in the context of the official statistical system. One section of the chapter deals with the degree to which measures of interpersonal and institutional trust are useful to inform policy making and the general public. Accuracy – and particularly validity – refers to the degree to which measures of trust capture the intended concept, which is a major concern for the measurement of trust. The chapter provides a brief overview of approaches to assessing validity, before considering both interpersonal and institutional trust in terms of the concepts of face validity, convergent validity and construct validity. Unlike most of the rest of the Guidelines, this section contains a substantial empirical component, which looks at the consistency of trust measures across different survey vehicles and over time.

2.2. A conceptual framework for measuring trust

To measure something, it is necessary to have a clear idea of the concept to be measured. Developing a sound conceptual framework is therefore an essential prerequisite to providing advice on the measurement of trust. A conceptual framework provides a working definition of the concept to be measured and sets out how the concept relates to other similar ones. In particular, a good conceptual framework should provide a clear definition of the concept to be measured and of any dimensions or components related to it.

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

This framework should be clear about what falls both inside and outside the scope for its measurement and relate the concept to specific measurement instruments.

In the case of trust, a sound conceptual framework is particularly important. Unlike some relatively straightforward concepts, such as age, gender or marital status, trust is inherently intangible. Although it is possible to observe trusting behaviour and to obtain self-reports from respondents about their stated levels of trust, it is not possible to directly observe trust as such. This raises the issue that respondents might not share a common view of what is meant with respect to trust in a survey question.

Theories of trust

There is an extensive literature on trust spanning a wide range of different disciplines within the social sciences, including political science, sociology, economics and psychology.[1]

In a review of what is known about measures of generalised trust, Nannestad (2008) identifies a wide range of definitions proposed by different authors who variously class the concept as a *general outlook on human nature*, an *affective attitude*, a *relationship*, a *decision* and an *action*. This highlights the degree to which there is not a single agreed definition of trust that can be adopted for these Guidelines. Nonetheless, several unifying themes cut across most measures of trust and provide a useful starting point. In particular, Hardin (2004) notes that trust is "in the cognitive category with knowledge and belief". This approach is echoed elsewhere: Morrone et al. (2009) note that expectations are central to most definitions of trust, and Uslaner (2008) identifies the central idea behind trust as a belief that most people share your moral values. A similar approach is shared by most other treatments of the subject, cutting across a wide range of different academic disciplines, from political science (e.g. Rothstein and Uslaner, 2005), to sociology (e.g. Delhey, Newton and Welzel, 2011) and economics (e.g. Zak and Knack, 2001).

A second common point shared by most theoretical approaches to trust is that they involve one person giving discretion to another party to affect the person's interests. There is thus an element of risk involved. This approach is spelled out explicitly by Hardin (2004), but is implicit in all accounts of the subject. Trust does not arise as an issue unless the person being trusted has the ability to affect the interests of the person doing the trusting materially or otherwise.

Beyond these points of commonality, however, there are significant differences in how trust is viewed. Nannestad (2008) identifies two main theories of trust: *rational trust* and *moralistic trust*. This distinction addresses the issue of whether trust should be thought of as a *belief* about other people's trustworthiness (rational trust) or as a *norm* regarding how to treat other people that is received from a person's parents, culture or environment (moralistic trust). In the latter case, it is assumed that trust is a belief about how a person should behave towards others (i.e. how much trust to extend to them) rather than a belief about how others are likely to behave towards you.

The strongest statement of rational trust is provided by Hardin (2004). Hardin defines trust in terms of a specific trusting relationship of the type **A** trusts **B** to do **X**, emphasising the strategic nature of trust. In particular, Hardin focuses on the notion of "encapsulated interest": in this formulation, trust occurs where the person or institution to be trusted has an incentive to be trustworthy because they internalise the interests of the person doing the trusting.[2] Their interest encapsulates the interest of the person trusting others. Because of the nature of this encapsulated interest – the person doing the trusting must

have good reason to believe that the person whom they trust has taken their interests on board – Hardin argues that trust is always "particularised". One must specify who is trusting, who is to be trusted, and what the person is to be trusted to do. In this sense, trust is strategic because it is conditional on a set of beliefs about the person being trusted and on their likely behaviour with respect to a given activity.

The main alternative to a rational notion of trust is the concept of moralistic trust. Uslaner (2008) argues that trust is inherited through socialisation rather than acquired. In this sense, trust is still an expectation about how others will behave, but it is not a strategic expectation. Trust is not grounded in the view that the person to be trusted encapsulates the interests of the person doing the trusting, but is rather a general attitude based on the life experiences and cultural background of the person doing the trusting. In Uslaner's formulation, trust is a "moral commandment to treat people as if they were trustworthy".

An important implication of the distinction between rational trust and moralistic trust lies in the role of reputation. In Hardin's formulation, trust becomes possible only in the existence of repeated interactions between the person doing the trusting and the person being trusted. Trust is rational in the sense that the trustor's expectation of good behaviour on the part of the person being trusted is based on either that person's reputation for good behaviour in the past or on a solid rational reason to expect that person's interests to align with those of the trustor. A restaurant owner, for example, might trust a supplier to provide the goods requested after being paid in advance because, if they are cheated by the supplier, then the supplier will lose all future business with the restaurant. The supplier's interests thus encapsulate the interests of the restaurant owner in this respect. This would appear to exclude an element of trust identified as important in much of the literature on the subject, i.e. the ability of trust to facilitate exchanges between people or to allow for people to co-ordinate actions in situations where they face each other in one-off situations in which their short-term incentives *do not* necessarily align. Many commercial and other exchanges between people are largely one-off and anonymous, effectively precluding trust as encapsulated interest, yet such exchanges still take place. In this perspective, Arrow (1972) stressed the importance of mutual confidence and trust in facilitating economic development.

In contrast, Uslaner's notion of moralistic trust allows trust to play a role in situations where the incentives of the trustor and of the person being trusted are not aligned. This approach is more in keeping with a view of trust that sees it as a mechanism for enabling strangers in a complex society to co-operate with each other. Fukuyama (1995) takes a similar view, emphasising the role of both repeated behaviour and shared social norms; in this approach, trust is understood as "the expectation that arises within a community of regular, honest and co-operative behaviour, based on commonly shared norms, on the part of the other members of that community."

Both rational and moralistic conceptions of trust define the concept as attitudinal, or in terms of the beliefs and expectations that people hold. An alternative definition of trust is provided by Fehr (2009), who argues for a behavioural definition of trust. Although Fehr notes that trust is tightly connected with preferences and beliefs, he argues that trust is best defined as a type of behaviour. In Fehr's approach, trust has two elements. The first element, which is unambiguously behavioural, is that the person who trusts places resources at the disposal of another party, without the means to guarantee that these will be returned. The second element is that the act of trust is associated with an expectation on the part of the person doing the trusting that the act of trust will be of direct benefit to them. This element

distinguishes trust from altruism and is associated with attitudes and beliefs rather than behaviour.

Conceptually, Fehr's behavioural notion of trust is compatible with either rational or moralistic trust. A behavioural view of trust abstracts entirely from the reasons why the person doing the trusting has an expectation of positive behaviour from the person who is trusted and focuses solely on the act of trusting itself. From the perspective of these Guidelines, this approach has the advantage that it does not pre-suppose a strong model of what creates trust. Instead, it treats trust as an empirical phenomenon that may be investigated independently of beliefs about how it is created.

Types of trust

All the theories of trust discussed above focus, in the first instance, on trust by people in other individuals. This is useful in establishing what is meant by trust conceptually. However, it is not sufficient to serve as the basis for measurement. There are obviously quite significant differences in asking about trust in another specific individual, trust in other people more generally or trust in a specific institution.

Table 2.1, below, provides one possible framework for thinking about how the different measures of trust fit together. This framework classifies measures of trust primarily in terms of the parties involved in the trusting relationship, and it has the advantage of capturing a very comprehensive range of situations. However, the framework also has some limitations in that it primarily focuses on distinguishing individual trust from the different elements of institutional and political trust.

Table 2.1. **A framework for multiple trust relationships**

By whom/on whom	Resident	Institution	Leaders
Resident	Interpersonal trust	Institutional trust	Political trust
Institutions	Civic trust	Inter-institutional trust	Political-administrative trust
Leaders	Political trust	Political-administrative trust	Multilateral trust

StatLink http://dx.doi.org/10.1787/888933584032

Despite its limitations, the framework in Table 2.1 is useful for narrowing down the scope of these Guidelines. Some of the types of trust identified by Table 2.1 are not suitable for measurement in surveys (e.g. civic trust) or do not obviously fall within the remit of official statistics (e.g. political trust), or both (e.g. multilateral trust). Trust by institutions or organisations (as opposed to trust in institutions) is not suitable for measurement through surveys of the general population, and hence none of the measures of trust by institutions are covered by these Guidelines. Similarly, a household survey cannot specifically target political leaders as respondents, making this group out of scope even if there were no other reason to avoid collecting data of this sort. However, collecting information from political leaders is clearly a significant problem from the perspective of maintaining the independence of national statistical offices. For this reason, the scope of the Guidelines is kept relatively narrow, focusing on people's trust in individuals (i.e. ***interpersonal trust***) and people's trust in institutions (i.e. ***institutional trust)***.

Some important distinctions highlighted by the literature on trust between different types of interpersonal trust are not well captured by Table 2.1. For example, Delhey et al. (2011) argue that, when framing survey questions, it is important to specify the relationship

between the person to be trusted and the person doing the trusting. For example, the level of people's trust in family or close friends might be expected to be different (narrow trust) compared to their trust in casual acquaintances or strangers (generalised trust). Delhey et al. also note that there are important differences across countries in the sort of relationship that people envisage when asked a generic question on trust in others. The degree to which these differences are important empirically will be explored later in the section of this paper on validity.

A similar set of issues to those raised by Delhey et al. (2011) can also be considered in the case of institutional trust. In particular, the specific institution that is being considered will have enormous implications for the degree of trust that people are ready to place in it.[3] While it is conceptually possible to consider institutional trust as a single construct, trust in specific institutions such as the police, courts, parliament or banks is likely to be of much greater practical relevance.

From a measurement perspective, the issue of people's trust in precisely which institutions or people should be measured is an important one. Obviously, users' needs provide an important guide as to which institutions or persons are relevant from a measurement perspective, but there are also empirical considerations. Respondents may not always clearly understand the differences between closely related institutions (e.g. between the legislative and the executive branch of government) or, even if they do, they may not have distinct views on all the institutions involved. Similarly, it matters a great deal whether responses to conceptually overlapping questions, such as those on trust in people you know personally as opposed to in neighbours or strangers, differ from each other. Evidence from the World Values Survey (WVS), which contains a wide range of questions on trust in different groups of people and institutions, suggests that, while there are differences in trust between different institutions and groups, there may be little practical value in making overly fine distinctions in trust questions (Box 2.1).

Box 2.1. **How many types of trust should be measured?**

The academic literature on trust includes a wide range of different measures capturing a diverse range of types of trust. Distinctions are made between different parts of government (e.g. parliament, the civil service, political parties, the current government) and between different types of persons (e.g. neighbours, personal acquaintances, family, people you meet for the first time). Many authors argue that such distinctions are fundamental to understanding trust meaningfully (e.g. Delhey, Newton and Welzel, 2011; Hardin, 2004). Such a view has obvious intuitive plausibility, as it is easy to imagine that trust in your family might be different from trust in someone you do not know personally. In practice, however, the main issue is whether the differences between these concepts are significant empirically. If respondents provide essentially the same answer to questions on trust in other family members and on trust in people they do not know personally, then there is little reason to collect information on the two concepts separately.

The World Values Survey provides a useful database for examining the degree to which different measures of trust capture different constructs. Wave 6 of the WVS contains questions on 18 different types of trust, with responses from 68 486 individuals, which allows an empirical analysis of how different responses group together. Table 2.2 shows the results of a principal component analysis of the WVS trust questions, aimed at identifying the degree to which the different questions collect fundamentally different information.

Box 2.1. **How many types of trust should be measured?** *(cont.)*

The table takes the 18 trust questions from wave 6 of the WVS and identifies the degree to which these reflect a smaller number of "latent" factors that best explain the variance in the dataset. For each trust measure, the higher the number in a particular column, the more strongly that measure is correlated with the latent factor associated with that column. Factor 2, for example, has high values for the government, political parties, parliament and the civil service, suggesting that the latent factor is related to political institutions. With a value of 0.4185 for factor 2, trust in the courts is clearly associated with trust in government institutions more broadly, to a much greater degree than trust in the armed forces (0.1342), but still not to the same degree as the core legislative and executive government institutions.

Table 2.2. **The dimensionality of trust measures in the World Values Survey**

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
The government (in your nation's capital)	0.1752	**0.7480**	0.0335	0.3271	0.0597
Political parties	0.2352	**0.8235**	0.0803	0.1339	0.0695
Parliament	0.2604	**0.8302**	0.0703	0.1655	0.0550
The civil service	0.3710	**0.6175**	0.0795	0.2346	0.0994
The armed forces	0.2348	0.1342	0.0597	**0.7278**	0.0983
The police	0.1687	0.3064	0.0662	**0.7708**	0.1152
The courts	0.2387	0.4185	0.0600	**0.6510**	0.0643
Major companies	**0.7160**	0.2356	0.0381	0.1722	0.0639
Banks	**0.6384**	0.2744	-0.0369	0.2069	0.0878
Universities	**0.6355**	0.2135	0.0764	0.2474	0.0470
Environmental organisations	**0.8024**	0.1589	0.0919	0.0753	0.0569
Women's organisations	**0.7783**	0.1101	0.1463	0.0629	0.0437
Your family	0.1084	0.0336	-0.0637	-0.0120	**0.7179**
Your neighbourhood	0.0415	0.1176	0.1914	0.1260	**0.7572**
People you know personally	0.0461	0.0336	0.3150	0.1434	**0.6838**
People you meet for the first time	0.0250	0.1079	**0.6301**	0.0688	0.3439
People of another religion	0.1019	0.0426	**0.8892**	0.0437	0.0549
People of another nationality	0.1230	0.0502	**0.8862**	0.0437	0.0610

Note: Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalisation. Total variance explained: 65.3%. Cronbach's Alpha: 0.8844.
Source: OECD calculations based on wave 6 of the World Values Survey (database) *www.worldvaluessurvey.org/WVSDocumentationWV6.jsp* as contained in the OECD Trust Database

StatLink *http://dx.doi.org/10.1787/888933584051*

An analysis of Table 2.2 shows that 5 different factors account for 65% of the total variance (across individuals) of the 18 trust questions included in the WVS. These factors are: trust in non-governmental institutions (factor 1: major companies, banks, universities, environmental organisations, women's organisations); trust in political institutions (factor 2: government of the day, political parties, parliament, civil service); trust in law and order institutions (factor 4: armed forces, police, courts); trust in people known to the respondent (factor 5: family, neighbours, people you know); and trust in strangers (factor 3: people you meet for the first time, people of another nationality, people of another religion). These broad results are robust to different specifications such as looking only at institutional trust or only at interpersonal trust, and hold across different waves of the WVS.

Similar factor analyses have been performed by other researchers. Naef and Schupp (2009) focus on a three-factor model using the German Socio-Economic Panel and find that interpersonal trust has two distinct factors relating to people you know and to strangers,

Box 2.1. **How many types of trust should be measured?** *(cont.)*

respectively, while institutional trust forms a distinct third factor. Uslaner (2002) finds similar results, but adds the important detail that the standard Rosenberg question on generalised trust (Rosenberg, 1957; also see below) is associated with the factor relating to strangers rather than people you know. Schneider (2016), who undertakes a more focused analysis of 35 countries in Eastern Europe and the former Soviet Union to examine the dimensionality of institutional trust, finds a similar split to that identified in Table 2.2 between trust in political institutions and trust in law and order institutions, although she does not test for non-governmental organisations.

This analysis confirms the salience of the conceptual distinction between interpersonal trust and institutional trust and identifies distinct sub-dimensions within each of these two main categories. However, it also highlights that many of the finer distinctions often made between different categories of trust are not very informative empirically. This does not necessarily imply that there is no value in asking more specific questions about trust: different trust questions may closely co-vary for reasons other than that they are measuring the same thing (e.g. they may have very similar drivers). If users have a sufficiently strong need, even relatively minor differences between closely related concepts may be important. However, it does suggest that, when deciding which trust measures are worth collecting, a relatively narrow range of measures will cover the most important aspects.

A definition of trust

To provide guidance on the measurement of trust, it is important to move from a description of different types of trust to a more specific definition that can provide a focus for the Guidelines. As discussed above, these Guidelines focus on trust ***by individuals in other individuals*** and ***in institutions***. A broad definition of trust is adopted, one that is clear and intuitive and that can serve as the basis for breaking down the broader notion of trust into more specific categories. For the purposes of the Guidelines, trust is defined as:

> a person's belief that another person or institution will act consistently with their expectations of positive behaviour.

This definition has two key elements. First, it focuses on a person's beliefs or expectations about behaviour. This is consistent with all accounts of trust found in the academic literature. Second, it focuses on expectations of positive behaviour. In other words, trust implies not only that a person being trusted will act consistently with expectations, but that those expectations are of good behaviour. Although it might be possible to state, in casual usage of the term, that a known thief can be trusted to take advantage of an opportunity to steal something, this is clearly not the concept that the Guidelines are trying to measure.

Among the theories of trust outlined earlier, the definition used here is closest to that proposed by Fehr, in that it takes no position on the way in which trust is formed. The key difference between the definition adopted here and Fehr's purely behavioural definition is that this definition focuses on ***belief*** rather than ***behaviour,*** which allows for measurement in situations where behaviour cannot be observed. The definition can also be located in the framework set out in Table 2.1. It implies that, for the purposes of the Guidelines, trust is to be understood as trust ***by individuals***, rather than by leaders or institutions. Conversely, the object of trust is wider, encompassing both other individuals and institutions. This reflects the fact that interpersonal trust and institutional trust are both policy-relevant.

In addition to distinguishing between interpersonal and institutional trust, it is also important to be precise about the specific categories of trust to be measured. Building on the factor analysis in Box 2.1, two categories of interpersonal trust as well as three types of institutional trust are identified. For interpersonal trust, these Guidelines distinguish between *generalised trust* and *limited trust*. Generalised trust refers to trust in people who are not known to the respondent or to trust in situations where the person being trusted is not specified. This is a concept of trust measured by the so-called Rosenberg question, first introduced in 1957, which asks, "generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" This can be contrasted with limited trust, which focuses on persons known to the respondent, including family, friends and people living in the respondent's immediate neighbourhood. Throughout the rest of these Guidelines, the term interpersonal trust will be used in situations that apply jointly to both generalised and limited trust, while the narrower terms will be used where there is a need to be more specific. As will be discussed later, generalised trust will be the primary focus for the Guidelines, due to its greater policy relevance and stronger evidence linking it to other outcomes of interest.

Institutional trust is used to refer to all types of trust in institutions, with trust in *political institutions*, *law and order institutions* and *non-governmental institutions* used to refer to a narrower concept. Although there is a high level of policy interest in trust in specific institutions, there is relatively little empirical evidence suggesting that respondents are able to make meaningful distinctions between these categories. For this reason, the Guidelines will make less use of these three categories than the distinction between generalised and limited trust.

In addition to the distinction between different types of institutions, there is a second important conceptual distinction to make with respect to institutional trust. A common theme in the academic literature on institutional trust (e.g. Nooteboom, 2007) is the distinction between "trust in competence" (i.e. whether the functioning of institutions matches people's expectations about the competencies of those steering them) and "trust in intentions" (which captures whether institutions act in a way that is perceived by people as ethical and fair). These distinctions are extended by Bouckaert (2012), who distinguishes between the "logic of consequences", where trust is derived causally from outcomes, and the "logic of appropriateness", where trust is based on the values of integrity and transparency. This distinction between the outcomes of an action and the intention that guided it forms the basis of the OECD trust framework, endorsed by the Organisation's Public Governance Committee. The framework identifies two key components of trust in institutions: *Competence* and *Values*. Within each component, relevant dimensions that are amenable to policy change are identified based on the common threads in the literature (Mcknight, Choudhury and Kacmar, 2002) and on the OECD update of this evidence (OECD, 2017a). The five dimensions identified are: *Responsiveness, Reliability, Integrity, Openness* and *Fairness*.

2.3. Approaches to measuring trust

Defining trust is a necessary precondition for measuring it, but it is not sufficient. To provide meaningful guidance on the best way of measuring trust, it is necessary to understand the various approaches to its measurement.

There is no single framework for classifying the different approaches to measuring trust in the academic literature. Nonetheless, drawing on a range of sources, it is possible to identify several distinct measurement approaches. At the most basic level, a long tradition of survey questions has directly asked people about their trust in others (e.g. Almond and

Verba, 1963) and institutions (e.g. WVS). Morrone, Tontoranelli and Ranuzzi (2009) make a distinction between traditional trust questions and attempts to measure trust through peoples' expectations about the behaviour of others (specifically, a survey question on whether a lost wallet is likely to be returned). Although the use of such "expectations questions", drawing on specific hypothetical scenarios is, so far, relatively limited, they set a distinctly different conceptual task for respondents than direct questions about trust, and thus might represent a source of additional information.

Beyond this, a large literature has compared actual trusting behaviour in experimental settings with survey questions on trust (Glaeser et al., 2000; Fehr et al., 2003; Gachter et al., 2004; Lazzarini et al., 2004; Naef and Schupp, 2009; Johnson and Mislin, 2012; Algan and Cahuc, 2013; Falk et al., 2016). Although experimental approaches to measuring trust might be regarded as beyond the scope of official statistics, these measures provide important insight into the validity of more conventional survey-based measures, and they have been used as a basis to develop better survey questions (Falk et al., 2015). Finally, it is important to note the existence of a wide suite of questions grounded in people's experiences that, while not focused directly on trust, can provide information on the subject. Taken together, these different approaches to measuring trust can be organised into four broad groups:

- evaluations
- expectations
- experiences
- experiments.

The different approaches to trust identified by these four groups do not represent substantively different concepts of trust, in the same way that, for example, *evaluative, affective* and *eudaimonic* measures capture substantially different constructs of subjective well-being (OECD, 2013a). Rather, they correspond to different approaches to measuring the same concept. The first three approaches are all survey-based and vary in the focus of the question asked to respondents. In particular, while evaluative questions focus on the respondent's views at the time of the interview, expectations questions have a future orientation, and questions on experiences focus on the past. In contrast, an experimental approach to measuring trust captures actual behaviour from respondents in a situation that is designed to elicit trusting (or non-trusting) behaviour.

Evaluations

Evaluative approaches to measuring trust focus on the respondent's response to questions on whether they have trust in an individual or institution. These questions ask the respondent to make an evaluation of their own feelings and/or beliefs at the current point in time, rather than recalling information about past experiences or speculating about the future. Most of the commonly used survey questions related to trust are evaluative. An example of an evaluative question is the standard generalised trust question by Rosenberg that is used in the WVS.

> "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"

Expectations

An alternative to asking for people's evaluation of how they feel is to ask them about their expectations of what would or will happen in a given, hypothetical situation. These

expectations questions are occasionally described as situational. Because these questions focus on expectations of how people will behave, it is sometimes argued that they are more specific and behavioural than is the case for evaluative questions (Morrone, Tontoranelli and Ranuzzi, 2009). However, the hypothetical nature of the question is also a potential weakness, as respondents may have difficulty providing an accurate answer to a hypothetical situation. The *lost wallet* question used in the Gallup World Poll is an example of an expectations question:

> "If you lost a wallet or a purse that contained items of great value to you, and it was found by a stranger, do you think it would be returned with its contents, or not?"

Experiences

An alternative to asking about what people expect to happen in a given situation is to ask the respondent about past experiences. Such questions do not focus on what the respondent feels or believes but rather on concrete events that have already taken place. This approach is the one customarily used for victimisation surveys, which often include questions related to experiences that represent low levels of trust, such as how often the respondent is fearful in their local town or city centre (e.g. New Zealand Crime and Safety Survey, 2014). Similarly, in the context of institutional trust, questions relating to people's experiences of corruption or discrimination have been used in both research and official surveys. Experience questions have not traditionally been a major source of data on trust, but are included here as they have the potential to provide information on the degree to which respondents have experienced situations where their expectations of (positive) behaviour from others have or have not been met. An example of this is the *experience of discrimination* question, which reflects a violation of social norms relating to the expectation of fair treatment. For instance, the German Satisfaction with Government Services Survey (Zufriendenheitsbefragung, 2015) asks respondents to assess their experience with a range of individual public agencies, including whether the respondent trusts the agency and felt discriminated against during the interaction with the institution. Similarly, the New Zealand General Social Survey includes the following question on the experience of discrimination:

> "The next question is about discrimination in New Zealand. By discrimination I mean being treated unfairly or differently compared to other people. In the last 12 months have you been discriminated against?"

Experiments

An alternative to collecting information on trust through surveys is to use experimental techniques to measure trusting and trustworthy[4] behaviour by participants, either in the field or in controlled laboratory conditions. Pioneered by experimental and behavioural economists and more recently by psychologists and neuroscientists interested in the cognitive processes underlying decision making, games such as the Trust Game (Berg et al., 1995)[5] collect information on the respondent's actual behaviour in circumstances where genuine rewards – typically, but not always, relatively small (yet large enough to form genuine incentives) – are at stake (Box 2.2). Because they measure actual behaviour, experimental methods avoid many of the limitations of survey-based approaches to measuring trust, such as social desirability bias in responses, or different understandings of the survey question between respondents.

Box 2.2. **Experiments for measuring trust**

Experiments can either take place in the real world or in the controlled setting of a laboratory, which is typically a fixed physical or virtual infrastructure that participants visit to undertake carefully scripted decision-making tasks. One approach to experimentally measuring trust in the field is illustrated by the 1996 experiment by the *Reader's Digest* (Felte, 2001). A number of wallets, each containing USD 50 in local currency and a fictitious ID, were "lost" in public places in a selection of major cities around the world. The proportion of wallets returned was then used by *Reader's Digest* to construct an index of trustworthiness for the different cities involved. This experiment was recently repeated in 2001, with over 1 100 wallets lost in different cities (Felte, 2001). The results were largely consistent with the picture provided by survey data. A high proportion of wallets were returned in the Nordic countries, New Zealand, Korea and Japan, while relatively few wallets were returned in Italy, Argentina and Mexico. A similar experiment was run in 2007 but with lost cell phones replacing wallets (Shanahan, 2007).

While the *Reader's Digest* field experiments with wallets and cell phones are intuitively appealing, they have a number of significant limitations. In particular, the number of wallets/cell phones involved in the experiments is low (10 wallets per city), and it is difficult to be confident that the proportion of lost objects returned is not influenced by differences in the particular areas where the objects were abandoned in the different cities. Although the *Reader's Digest* experiments have been influential, leading the Gallup World Poll to use a lost wallet question (see above, under "Expectations"), the significant costs and methodological limitations associated with "losing" large numbers of wallets in different cities mean that this approach has never become a major source of trust data.

In the laboratory, a more systematic and rigorous approach to testing trust and trusting behaviour experimentally was developed by Berg, Dickhaut and McCabe (1995). The Trust Game has since become the standard approach to measuring trust experimentally in controlled settings, and it has formed the basis for a wide range of different variant games testing different aspects of trust and trusting behaviour. The game has now been repeated in dozens of countries involving tens of thousands of individuals, allowing for large-scale meta-analyses of the results (e.g. Johnson and Mislin, 2011; Johnson and Mislin, 2012; Zelmer, 2003). Although conducted under controlled conditions, and typically for rather small stakes, the main findings of the Trust Game have been replicated with relatively high stakes, and have been found not to be affected by social desirability bias or by the range of strategies allowed to participants (Naef and Schupp, 2009). There is today a growing body of literature examining the relationship between laboratory results from the Trust Game and outcomes in the real world.

The original Trust Game developed by Berg et al. involved two anonymous players in different rooms of a laboratory. Each player starts with an endowment of USD 10 of real money provided by the experimenters. Player 1 is then invited to give some (or all) of their endowment to player 2. Any amount given by player 1 is tripled by the experimenters so that if, for example, player 1 gives USD 5 to player 2, the second player will receive USD 15. After the first transaction, player 2 is then provided with the opportunity to return some (or all) of their money to player 1. This amount is not adjusted by the experimenters in any way. Given the structure of the experiment, the total payoff to both players is maximised when player 1 gives all his/her money to player 2, in which case the total payoff would be USD 40 (USD 10 from player 1, tripled to USD 30, plus the USD 10 original endowment of player 2); if player 2 is trustworthy, and sends half their money back to player 1 at the end of the game, this would allow each player to walk away with USD 20, as opposed to their original endowment of only USD 10 each. However, player 1 has no way to ensure that player 2 responds in this way so that, from the perspective of player 1, there is always a risk that any money given to player 2 is simply gone.

The Trust Game produces two measures of trust. First, the proportion of money given by player 1 to player 2 can be taken as a measure of trust by player 1 that player 2 will reciprocate any money sent to him or her and will respond in kind. Second, the proportion of money returned by player 2 to player 1 can be interpreted as a measure of trustworthiness

of player 2. Repeating the game over a representative sample of the population allows estimating the level of trust and trustworthiness among participants.

However, experimental games also have limitations and might be less relevant when the goal is to measure trust among the wider population. Lab games can be expensive to run, and most laboratory studies have so far been restricted to university students, thus basing their findings on small samples that, apart from being unrepresentative of the whole population, can be affected by self-selection bias.[6] This is a significant limit when it comes to establishing the external validity of the findings of many laboratory experiments. Algan and Cahuc (2013) review a range of studies indicating that experimental behaviour in a laboratory setting is more effective than survey responses in predicting a range of real-world outcomes. Further encouraging evidence on the external validity of lab experiments comes from Naef and Schupp's (2009) finding that both survey and experimental measures of trust correlate with related factors such as being an entrepreneur or a shareholder. Being an entrepreneur increases exposure to risk, particularly to the risks associated with informal agreements with other people (Guiso et al., 2006), while investing in the stock market implies trust in both the stock market system as a whole and in the executive management of the companies being invested in (Guiso et al., 2008). Furthermore, Karlan (2005), in a study of borrowers in a Peruvian microcredit programme, shows that behaviour in a Trust Game predicted repayment rates of the participants' loans one year later.[7]

While experimental techniques such as the Trust Game may not be suitable for direct inclusion in most large-scale sample surveys run by national statistical offices, they are relevant to these Guidelines in two ways. First, experimental data provide one of the best sources of information for testing the validity of different survey measures and for developing better survey questions (Fehr et al., 2003; Naef and Schupp, 2009; Falk et al., 2015). This is important given the challenges involved in testing the validity of measures of intangible concepts such as trust. Second, although national statistical offices may not be well positioned to collect experimental data in large sample surveys, this does not mean that such an approach cannot be pursued. The *Trustlab* project aims to do exactly this for a number of OECD countries: over time, the evidence gathered through this project will provide a rich base of information on trust collected on a comparable basis from a representative population sample across the OECD (Box 2.4).

2.4. The statistical quality of trust measures

For measures of trust to be taken seriously, statistical quality is crucial. Unless data accurately capture the concept being measured, there is little point in collecting it. This is particularly true for official statistics, which are expected to be of the highest quality. The United Nations Fundamental Principles of Official Statistics state that "official statistics provide an indispensable element in the information system of a society, serving the government, the economy and the public with data about the economic, demographic, social and environmental situation" (United Nations, 2014). It is therefore crucial that these Guidelines are solidly grounded in a clear understanding of the quality of trust measures.

The OECD's approach to dealing with issues of statistical quality is set out in the *Quality Framework and Guidelines for OECD Statistical Activities* (OECD, 2008). The Framework defines quality as *fitness for use* in terms of user needs. The ultimate benchmark of the quality of statistics is whether they meet the needs of the user in terms of providing useful information. Because users must often make decisions about a course of action, whether or not statistical information is available, it is helpful to focus on the *fitness for purpose* of

measures rather than on whether the measure in question provides a "perfect" representation of the concept that it is intended to capture. This may involve accepting the use of data that are less than perfectly accurate, provided that they are of sufficient quality to improve on rather than detract from the quality of decision making.

The OECD framework identifies seven dimensions of statistical quality. These seven dimensions define the characteristics of high-quality data and provide a structured way of assessing the quality of a particular set of statistics. These dimensions are:

- *Relevance*, i.e. the degree to which data address the purposes for which they are sought by users.
- *Accuracy*, i.e. the degree to which data correctly describe the quantities or characteristics that they are designed to measure.
- *Credibility*, i.e. the confidence that users place in statistics based on the reputation of the data producer.
- *Timeliness*, i.e. the length of time between the availability of data and the phenomenon or event that the data describe.
- *Accessibility*, i.e. how readily the data can be located and retrieved by users.
- *Interpretability*, i.e. the ease with which users can understand and properly use and analyse the data.
- *Coherence*, i.e. the degree to which the data are mutually consistent with other similar measures and logically integrated into a system of statistics.

These seven criteria, along with the more general principle of cost effectiveness in producing/collecting such data, define the OECD's overall framework for assessing statistical quality. However, most of these criteria relate to ***how*** statistics are measured and collected rather than to ***what*** is collected. Credibility, for example, is a quality of the data producer, rather than of the measure. Similarly, timeliness, availability, interpretability and coherence largely relate to the broader context in which data are collected and made available to the public, rather than to the choice of measures used. For the purposes of these Guidelines, the concern of statistical quality is more narrowly focused on what should be collected and whether the proposed measurement instruments are fit for purpose. Thus, the main focus for assessing the quality of measures of trust will be the principles of relevance and accuracy.

Relevance

Data are relevant when they meet the needs of users. In the case of official statistics, where there may be a legal requirement for respondents to participate in data collection, it is of particular importance that information is collected only where there is a clear use for the data and where there are no obvious alternative information sources. For this reason, it is essential to identify the purposes of trust data before developing advice on measurement. Clarity about the relevance of trust data helps both to identify which trust measures are important to collect and to provide assurance that trust measures are worth collecting in the first place.

Modern societies are increasingly complex, and ill-informed policy choices may incur high costs and be difficult to reverse. Information about how key societal outcomes are changing, the drivers of those changes, and the impact of policy on these outcomes is therefore essential to good decision making. The ability of data to inform policy is thus central to relevance, particularly for national statistical offices. Beyond this, there is a broader public interest in accessing information on key social and economic outcomes.

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

This interest is distinct from that of policy making and is grounded in the value of an informed public. Trust data can therefore be considered to be relevant if they inform policy or the wider public.

Two core uses of trust data can be identified. These are:

- monitoring progress (current levels of people's well-being and sustainable development for the future)
- evaluating the effectiveness of government policies.

These two core uses of trust data are of relevance both from the perspective of policy making and from that of the general public. Policy makers have a clear interest in being able to evaluate whether their decisions are resulting in the desired outcomes and in understanding how effective different government institutions are at implementing policy. However, assessing sustainability, well-being and the effectiveness of government are also of interest to the general public from the perspective of democratic accountability.

Monitoring progress (current levels of people's well-being and sustainable development for the future)

The past decade has witnessed an increasing consensus on the need to measure the outcomes that are the ultimate goals of policy (well-being) and to account for all of the capital stocks that contribute to the sustainability of those outcomes over time. This two-part model, focusing on flows of population well-being underpinned by stocks of natural, economic, human and social capital, is at the core of the modern consensus on the measurement of well-being and of sustainable development (Stiglitz, Sen, and Fitoussi, 2009; OECD, 2011; UNECE, 2014). Central to this approach is a multidimensional view of people's well-being, seen as encompassing all those different outcomes – objective and subjective – that are valued by people in and of themselves or that are fundamental to allowing people to pursue the sorts of lives they have reason to value. This perspective is complemented by the recognition that measuring changes in the capital stocks used to produce well-being is essential to understanding sustainability. This perspective is reflected in reports by international organisations such as the OECD's *How's Life?* (2011, 2013c, 2015) and, most recently, in the United Nations Sustainable Development Goals (SDGs) (UN, 2015; OECD, 2017b). National initiatives to measure well-being and sustainable development also reflect this general framework, including those pursued in the United Kingdom (ONS, 2016), Israel (Ministry for Environmental Protection, 2014), Italy (ISTAT, 2013) and many other countries. Both of the two dimensions of trust covered by these Guidelines – interpersonal and institutional trust – are of vital importance to measuring well-being and sustainability.

The United Nations Sustainable Development Goals (SDGs)

A major task for national statistical offices over the next decade will be measuring progress towards the SDGs. These goals set out an ambitious plan of action for *people, planet, prosperity* and *peace*, to be achieved through *partnership*, and with the overarching objective of leaving no one behind. At the core of this plan are 17 SDGs comprising 169 targets. The 2030 Agenda set out by the United Nations (UN, 2015) allows countries to take their own approach to measuring progress towards the SDGs, but "encourage[s] member states to conduct regular and inclusive reviews of progress at the national and sub-national levels". While this resolution allows for some discretion in how to measure progress, it pushes nations very strongly towards measuring the SDGs in some kind of way.

Trust – both institutional and generalised trust – is central to monitoring progress towards the SDGs. In particular, Goal 16 of the SDGs ("Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels") focuses explicitly on governance, institutional quality and political participation. These themes have clear links to trust, and this is recognised in the work of the Praia Group on Governance Statistics. The explicit aim of the Praia Group is to improve the measurement of governance and encourage the production of internationally comparable statistics on governance with a view to informing Goal 16 of the SDGs.[8] This will involve a work programme covering different elements of governance, including citizens' trust in public institutions, their experiences of the performance of public institutions, and measures of generalised trust.

The Praia Group's deliberation of institutional trust as a potential indicator for Goal 16 of the SDGs is based primarily on its importance as a direct measure of institutional quality (Klijn, Edelenbos and Steijn, 2010). Institutional trust is identified in the OECD's *How's Life?* report (OECD, 2011) as an important aspect of civic engagement and governance. In discussing the statistical agenda ahead for this dimension, the OECD (2011) identified a number of actions aimed at improving measurement in this area:

> …better measures are required to assess how people perceive the quality of the democratic institutions in the country where they live. While many unofficial surveys contain questions for measuring civic engagement and governance, they typically have a narrow geographic coverage, small sample size and inadequate sampling procedures. Steps should be taken to include questions on civic participation and trust in institutions in large-scale official surveys, through a combination of a few recurrent questions in regular surveys and more detailed questions in (less frequent) dedicated modules, as is already done in some OECD countries (e.g. special modules of the Current Population Survey on voting and civic engagement in the United States).

The OECD decision to use measures of institutional trust among the indicators of the civic engagement and governance dimension of well-being is supported by analyses of the drivers of subjective well-being. Boarini et al. (2012) find that confidence in the judicial system has a significant positive impact both on people's satisfaction with their own life and on their affect balance, after controlling for a wide range of other potential drivers of well-being.[9] With respect to the effect on life satisfaction, the magnitude of this impact is reported as 1.1 times the impact that a doubling of individual income would have. Similarly, Helliwell and Putnam (2004) find a robust association between trust in the police and life satisfaction, with similar results replicated with different datasets in Helliwell and Wang (2010).

Generalised trust is also highly relevant as a measure of the quality of governance (OECD, 2011). This is because there are feedback loops from institutional quality to levels of generalised trust in society. Rothstein and Uslaner (2005) argue that high levels of generalised trust depend crucially on the effective functioning of governance and justice institutions. It is only when these institutions are performing both fairly and competently that it is possible for a person to extend trust to strangers without putting themselves at risk. Furthermore, there is also a causal pathway running in the opposite direction, i.e. from high levels of generalised trust to fair and effective public institutions. In particular, public institutions are likely to be fairer and more effective in societies where individual civil servants have reason to believe that other officials and citizens are trustworthy, and where the cost of corrupt or unfair behaviour is likely to be high and the risk of being

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

caught out in such behaviour is more probable. For this reason, even though generalised trust itself is conceptually distinct from institutional trust and from governance more generally, it is nonetheless a powerful indicator of the quality of governance.

Measuring social capital for sustainable development

In addition to contributing to monitoring Goal 16 of the SDGs, measures of trust are of importance for monitoring the stocks of capital that underpin the production of well-being. The OECD and the UNECE (UNECE, 2014) define sustainability as a situation in which current levels of well-being occur with stable or increasing capital stocks. This implies the need for adequate measures of the capital stocks underpinning sustainable development – economic capital, natural capital, human capital and social capital (OECD, 2011). Social capital is used here in a narrower sense than in some of the wider literature. The term social capital has been applied to a number of different concepts, including the networks and values that enable elites to transmit social advantage across generations (Bourdieu, 1984), the resources inherent in networks that allow people to pursue individual outcomes (Coleman, 1988), and socially useful norms and values (Putnam, 1993). But it is only the latter concept that is relevant to the measurement of sustainable development. The OECD uses a variant of this definition when it describes social capital as the "shared norms, values and understandings that facilitate co-operation within or among groups" (OECD, 2001).

Although social capital is important, as an intangible capital stock it is difficult to measure. Residual-based estimates of the importance of intangible capitals (i.e. human and social capital) suggest that the magnitude of the effect of these intangible stocks on other outcomes is large (World Bank, 2006; Hamilton and Liu, 2013). In order to move beyond the general observation that social capital is important in producing other valued outcomes to useful policy advice, it is essential to have direct measures of the size of the social capital stock and of how it changes over time. Generalised trust is one of the best available proxy measures of social capital, and better measures of generalised trust would make a very significant contribution to understanding social capital, its drivers and consequences (Scrivens and Smith, 2013). The role of generalised trust in measuring social capital is also stressed by other authors (e.g. Knack and Keefer, 1997; Helliwell and Putnam, 2004; Morrone, Tontoranelli and Ranuzzi, 2009; Sturgis and Smith, 2010; Algan and Cahuc, 2013). However, for generalised trust to inform policy, adequate information must be available. Existing analyses are largely static and cross-sectional in nature, and provide only limited information on trends in stocks of social capital or on the impact of policy changes on this stock. Key limitations of existing data are the relatively small sample size of existing non-official surveys such as the WVS and the lack of sufficiently regular data to enable meaningful analysis of changes in interpersonal trust over time.

Indicators of well-being

Interpersonal trust also has an obvious relevance as an indicator of well-being. It is unpleasant to live in a situation where a person feels that he/she cannot trust the people around them, and low levels of trust in others place important limits on how a person lives their life. If well-being is viewed in terms of the capabilities a person has to live the sort of life they have reason to value (Sen, 1992 and 2009), then trust in other members of society is in itself an important element of well-being. In particular, the level of trust a person has in his/her fellow citizens shapes the sort of choices that a person might feel comfortable making (Alsop et al., 2006). With this rationale in mind, the OECD has chosen generalised

trust as one indicator of its "social connections" dimension in *How's Life?* (OECD, 2011). Similarly, from a more utilitarian perspective on well-being, interpersonal trust is one of the main drivers of subjective well-being.

The importance of interpersonal trust for assessing well-being is highlighted by the *World Happiness Report* (Helliwell et al., 2016), which presents an analysis of the main drivers of country differences in subjective well-being across the globe. Although interpersonal trust itself is not included in the analysis, as it is not part of the dataset on which the *World Happiness Report* is based (the Gallup World Poll), access to social support – which is closely correlated with trust in others – is one of the six main determinants of cross-country differences in life satisfaction identified in the report.

More directly, using data from the Gallup World Poll and the Canadian General Social Survey, Helliwell and Wang (2010) investigate the relationship between generalised trust and subjective well-being in more depth. After controlling for income and a range of social and demographic factors, the authors find a robust relationship between generalised trust and people's life satisfaction; this positive relation holds regardless of whether trust is measured via the traditional (Rosenberg) question on generalised trust or via a question about whether a lost wallet would be returned. Boarini et al. (2012) use data from the Gallup World Poll to test whether the dimensions of well-being used in the OECD Better Life Initiative correlate with overall life evaluations: they find that the average level of generalised trust has a significant positive impact on life satisfaction, after controlling for other dimensions of well-being. Fleche, Smith and Sorsa (2012) report similar results in a cross-country analysis using data from the WVS.

However, it is not only generalised trust that matters directly for well-being. There is also reason to believe that institutional trust plays an important role in well-being over and above its instrumental value in achieving other outcomes. While institutional trust is certainly important in this regard, the justification for measuring institutional trust here is its intrinsic importance to people's well-being. This is explored by Frey, Benz and Stutzer (2004) and Frey and Stutzer (2005, 2006), who highlight the importance to well-being of "procedural utility", i.e. the process through which important decisions are made. This approach suggests that procedural utility has an important impact on people's well-being, independently of the resulting outcome. In this perspective, the welfare gains from a policy decision that increases total income can be partially or totally offset by losses resulting from a decision-making process perceived by people as unfair or illegitimate. This approach strongly underscores the importance of institutional structures for people's well-being. The argument about the importance to people of being part of the process through which collective decisions are made and enforced is echoed by Alkire (2002), who surveys a wide range of international initiatives to define the key dimensions of "human flourishing", concluding that items relating to political participation and processes feature in nearly all of these.

Evaluating government effectiveness

While generalised trust is of primary importance for measuring social capital, institutional trust is most relevant to evaluating the effectiveness of government policies and programmes. Institutional trust is fundamental to the effective functioning of institutions and networks (e.g. Klijn, Edelenbos and Steijn, 2010). When people have a high level of trust in institutions, compliance is less of a problem, and it is easier to implement policies that may involve trade-offs between the short and long-term, or between different parts of society (Marien and Hooghe, 2011; OECD, 2013b). Institutional trust is especially important to

government activities that address market failures (e.g. public health or environmental protection) or where long-term gains require short-term sacrifices (e.g. education and pensions). In most of these cases, individuals cannot obtain an immediate, exclusive benefit from policy actions, and there is a high risk of free-riding. In these cases, trust in institutions may encourage people to accept some personal sacrifice in the short run for the sake of a larger benefit in the future, for themselves or for the community at large. Trust in institutions may be important also in cases, such as obtaining a business license, where decisions are influenced by the reliability of the rules that frame decisions or by the reliability of the institutions that make these rules. Thus, while high levels of institutional trust reduce government failures and the costs of enforcement and transaction, a lack of institutional trust provides incentives to opportunism, informality and free-riding, compromising the effectiveness of public policies.

A significant body of research has looked at the relationship between trust in institutions and government effectiveness. This research is based largely on unofficial data, but identifies some of the key elements in the relationship between institutional trust and government performance. One key theme here is the view that institutional trust depends on the congruence between people's (and businesses') preferences (i.e. their interpretation of what is right and fair) and their perceptions of how the government functions (Bouckaert and van de Walle, 2003). While a core theme cutting across this literature is the value of institutional trust as a measure of government effectiveness, this research also highlights the need for more detailed information. In particular, better measurement of institutional trust requires not only narrowing down the scope of the trust relationship to be measured, but also using a more operational notion of trust that can be deconstructed into meaningful policy questions.

Accuracy

Accuracy is of fundamental importance to any statistical measure. If a proposed measure does not correctly reflect the underlying concept that it is intended to capture, then it fails the basic test of providing useful information. Accuracy is typically assessed by looking at two dimensions of the measure, i.e. its ***reliability*** and ***validity***. The reliability captures the degree to which a measure produces consistent information over time and across different measurement vehicles: a measure that takes different numerical values when it is repeated over short time periods is clearly unreliable, and fails a basic test of accuracy. Reliability is thus concerned with the *variance* of the measure. Validity is concerned not with variance of the measure but rather with its *central tendency*, i.e. does the measure actually reflect the underlying concept to be measured? A measure can be valid but not reliable if it produces the correct result on average, but with a wide variance. Similarly, a measure can be reliable but invalid, when it captures items unrelated to the variable of interest.

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). Where validity cannot be directly assessed, a measure can be considered valid if it performs well in terms of all three aspects of validity outlined above.[10] Ideally, this would be assessed through a large and well-developed scientific literature that covers all the main types of validity, and which is sufficiently mature for a consensus to have emerged.

In some cases, a large and developed body of literature for a measurement concept will already exist. For example, the *OECD Guidelines on Subjective Well-being* were based on an extensive body of research by experimental psychologists on the validity of different types of subjective well-being measures. This reflected more than a decade of intense interest in the measurement of subjective well-being from a large number of highly regarded economists, psychologists and sociologists. Unfortunately, the same is not true with respect to the measurement of trust. While some excellent academic work on trust exists, taken as a whole, this body of work lacks the size, rigour and methodological focus that characterised work on the validity of subjective well-being measures. Because of this, a substantial focus for these Guidelines is to assess the validity of trust measures; this involves an empirical component to the work that was not present in the *OECD Guidelines on Measuring Subjective Well-being*. This empirical work, which is summarised in this chapter, is based largely on the OECD Trust Database (Box 2.3).

Box 2.3. **The OECD Trust Database**

A key challenge in building a better understanding of the drivers of trust and its impact on other outcomes is the limitations of the available data. This limitation has also had an important impact on the ability to assess the validity and reliability of trust data rigorously.

For a long time, the only substantial source of cross-country trust data was the World Values Survey (WVS). Recently, a wider range of non-official data have become available. The OECD Trust Database is an effort to map existing sources of (so far non-official, apart from EU-SILC) data and compile them into a single repository of information. Table 2.3 displays the different surveys included in the OECD Trust Database, which has been constructed as a cross-country panel dataset. Its coverage goes beyond OECD member states and includes up to 124 countries depending on the survey considered, spanning the period between 2002 and 2015 or the latest year available. The year 2002 was chosen as the database's starting point, since the number of surveys regularly collecting trust data doubled with the appearance of the European Social Survey (ESS) in that year.

Table 2.3. **Surveys included in the OECD Trust Database and their key characteristics**

Survey	Inception	Frequency	Number of countries in the OECD Trust Dataset	Coverage of the OECD Trust Dataset
EU Statistics on Income and Living Conditions (EU-SILC)	2003	2013 ad-hoc module	33	2013
Gallup World Poll (GWP)	2006	Yearly	115	2006-2015
World Value Survey (WVS)	1981	Every 5 years	6 (Wave 4) 46 (Wave 5) 45 (Wave 6)	1999-2004 (Wave 4) 2005-2009 (Wave 5) 2010-2014 (Wave 6)
European Social Survey (ESS)	2002	Every 2 years	22 (Round 1) 25 (Round 2) 23 (Round 3) 28 (Round 4) 27 (Round 5) 29 (Round 6)	2002 (Round 1) 2004 (Round 2) 2006 (Round 3) 2008 (Round 4) 2010 (Round 5) 2012 (Round 6)
European Quality of Life Survey (EQLS)	2003	Every 3 years	31 (Round 2) 34 (Round 3)	2007-2008 (Round 2) 2011-12 (Round 3)
Eurobarometer	1973	Yearly	34	2003-2015
Latinobarometer	1995	Yearly	19	2002-2015

StatLink http://dx.doi.org/10.1787/888933584070

Box 2.3. **The OECD Trust Database** *(cont.)*

Different surveys have different geographical coverages and collection frequencies. In the case of the Gallup World Poll, data are collected annually for countries in all regions of the world, while for other surveys – including the WVS, ESS and the European Quality of Life Survey (EQLS) – the coverage is limited to a smaller set of countries. Data collection takes place every 2 years for the ESS, every 3 years for the EQLS and roughly every 5 years for the WVS.

Both institutional and interpersonal trust feature in the OECD Trust Database. In the case of institutional trust, questions are traditionally formulated through a common heading (e.g. *do you have confidence in your…*) followed by a list of primarily public institutions (e.g. government, congress, etc.) and less commonly private (e.g. major companies). Survey wording varies considerably, both in terms of the general construction of the question and in the use of the term trust or one of its various synonyms (e.g. confidence). Also, different surveys, or even different questions within the same survey, make use of different response scales. For example, while the GWP Poll relies primarily on a *yes/no/don't know* response format, other surveys such as the ESS and EQLS use longer numeric scales (0-10 and 1-10 in the case of ESS and EQLS, respectively). In the case of the WVS, questions are usually answered using a 4-point Likert scale (i.e. *a great deal*, *quite a lot*, *not very much* and *none at all*). In the OECD Trust Database, different questions have been re-scaled to a binary *yes/no* format that allows comparability across surveys.

Further observation of the available data on institutional trust reveals that questions sometimes refer to similar concepts while using quite different descriptions. For instance, some surveys refer simply to *the courts* while others ask about the *judicial system*. While in most cases the interpretation of these concepts is straightforward, in others the lack of clarity may have more significant implications. For example, although most surveys ask about trust in government, the ESS asks about trust in politicians, and EU-SILC addresses trust in the political system. In turn, the Latinobarometer has included questions about trust in government, the state and public administration, concepts traditionally related to each other but not strictly synonymous. The EQLS includes a few questions worded in a more comprehensive way: instead of asking about a particular institution, the EQLS has extended the question to a system (e.g. state pension system or social benefit system) that includes a set of institutions.

Generally, non-official household surveys provide less information on interpersonal trust compared to institutional trust and, where interpersonal trust is included, the focus is limited to generalised trust. The most common question (asked by four of the surveys under study), which is very similar to the version introduced by Rosenberg in 1957, is the following:

"Generally speaking, would you say that most people can be trusted or you can't be too careful in dealing with other people?"

In the case of the WVS, in addition to *most people*, the questions inquire about additional parties to be trusted (e.g. family, neighbourhood, known people, people met for the first time, etc.). Only the EQLS asks a direct question about trusting people in general.

The surveys included in the OECD Trust Database vary in terms of data quality. While all of the surveys have different strengths and weaknesses, some of them place a greater emphasis on methodological rigour and consistency than do others. Both the ESS and EQLS are directly funded by the European Commission and aspire to very high standards of data quality. A great deal of attention is paid to consistency across countries, and changes between waves

Box 2.3. **The OECD Trust Database** *(cont.)*

are carefully managed. Covering a much greater range of countries, the Gallup World Poll varies more in survey content from wave to wave, but retains a strong focus on methodological consistency and minimises the impact of questionnaire changes by having a fixed core questionnaire. The WVS has evolved over time, and data quality is higher in more recent waves than in earlier waves. In contrast, both the Eurobarometer and Latinobarometer put more emphasis on responsiveness to policy issues, and thus the questionnaires change more frequently and the response rates are generally lower. The OECD Trust Database is available on-line as an electronic annex of Gonzalez and Smith (2017).

Reliability

As discussed above, the reliability of a measure captures the degree to which it produces consistent results when measured at different times. Ideally, a good measurement instrument should produce the same estimate of the concept measured whenever it is applied in similar circumstances, and should produce a different result only when there has been a substantive change in the object of measurement. 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. Unfortunately, there are relatively few studies that investigate the test-retest reliability of trust measures. One of the few studies that does is Naef and Schupp (2009), who examine the reliability of measures of generalised trust: the authors find a relatively low correlation coefficient between the two time periods at which trust was measured (0.48) but adequate levels of reliability when considering three measures of generalised trust jointly. Taken on its own, this appears to point towards a need for multi-item measures of trust in order to obtain sufficient levels of reliability. However, an alternative strategy for assessing the reliability of trust measures is available. The OECD Trust Database provides repeated measures of trust for a wide number of countries at different points in time, from different surveys, and using different measurement instruments. This provides a strong basis for assessing the reliability of trust measures at the country-average level. In particular, if different surveys produce consistent results for different countries, despite differences in timing and methodology, then it can be said that trust measures display a high degree of reliability at the cross-country level. Evidence of reliability based on analysis of data in the OECD Trust Database is presented below separately for institutional and interpersonal trust.

Institutional Trust. The wide range of questions on institutional trust available in the OECD Trust Database, and the fact that all surveys include at least some of them, makes it possible to conduct a fairly thorough analysis of the reliability of these measures. This includes looking at whether the reliability of questions on institutional trust is different for different types of institution. Table 2.4 reports the correlation coefficients between measures of trust in the national government for seven different surveys. In each case, the comparison is between country-average levels of trust in the national government for the same country and year, but as measured through different surveys. Because the surveys may be carried out at different periods within the same year, and because each survey has a unique sample of respondents, some difference in reported levels of trust may be expected. Still, a high correlation coefficient provides strong evidence that the measures produce reliable information.

Table 2.4. **Correlation in trust in the national government across surveys**

Survey	GWP	EB	ESS	WVS	EQLS	LB	EU-SILC
GWP							
EB	0.84						
ESS	0.81	0.74					
WVS	0.84	0.91	..				
EQLS	0.89	0.91	0.89	..			
LB	0.77	..	..	0.59	..		
EU-SILC	0.81	0.85	0.87	..	0.86	..	

Note: The Pearson correlation coefficient is a test for the strength of a relationship between two variables or datasets. Under this method, it is assumed that the data are normally distributed and that the expected relationship between them is linear. If the correlation between the two items is high, this suggests that the two measures capture the same underlying concept. GWP = Gallup World Poll; EB = Eurobarometer; ESS = European Social Survey; WVS = World Values Survey; EQLS = European Quality of Life Survey; EU-SILC= European Union Statistics on Income and Living Conditions; LB = Latinobarometer. The ESS question refers to trust in politicians rather than the government.
Source: Calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584089

Table 2.4 shows correlation coefficients between surveys of between 0.59 (Latinobarometer and the WVS) and 0.91 (Eurobarometer and both the WVS and the EQLS). With the exception of the Latinobarometer/WVS result, all the correlation coefficients are above 0.74, and the majority are well above 0.80. It is worth noting that all correlations below 0.80 involve either the Latinobarometer or the Eurobarometer, which are the two surveys included in the analysis with the lowest levels of methodological rigour. In particular, both surveys involve relatively frequent changes both to the questionnaire and to the time when the survey is in the field, and response rates are relatively low. It is also of interest that the European Social Survey, which asks about trust in politicians rather than in national governments, produces a lower correlation than other surveys.

Table 2.5 reports a similar analysis that focuses on trust in the judicial system rather than the national government. Again, the results obtained from different surveys correlate with each other very strongly, ranging from 0.68 (Latinobarometer/GWP) to 0.93 (European Social Survey with the GWP and with Eurobarometer). Ignoring the Latinobarometer, the range of correlations is even narrower (0.82 to 0.96) than in the case of trust in the national government, reflecting the fact that all of the surveys reviewed here include relatively similar

Table 2.5. **Correlation in trust in the judicial system across surveys**

Survey	GWP	EB	ESS	WVS	EQLS	LB	EU-SILC
GWP							
EB	0.90						
ESS	0.93	0.93					
WVS	0.82	0.82	..				
EQLS	0.92	0.84	0.90	..			
LB	0.68	..	..	..	..		
EU-SILC	0.92	0.96	0.95	..	0.91	..	

Note: The Pearson correlation coefficient is a test for the strength of a relationship between two variables or datasets. Under this method, it is assumed that the data are normally distributed and that the expected relationship between them is linear. If the correlation between the two items is high, this suggests that the two measures capture the same underlying concept. GWP = Gallup World Poll; EB = Eurobarometer; ESS = European Social Survey; WVS = World Values Survey; EQLS = European Quality of Life Survey; LB = Latinobarometer, EU-SILC= European Union Statistics on Income and Living Conditions.
Source: Calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584108

questions on trust in the judicial system. Similarly, when the same analysis is performed on trust in the police, the correlation coefficients range from 0.75 to 0.90.

Taken together, Tables 2.4 and 2.5 support the view that measures of institutional trust are reliable at the country-average level. For all instances where the question is directly comparable and the surveys are of relatively high quality, cross-country correlations are in excess of 0.80 and typically above 0.90 for those surveys with the highest data quality (the GWP, the ESS and the EQLS). Even where the data quality is lower, the correlation between country averages is still respectable.

Interpersonal Trust. Evidence from the OECD Trust Database on the reliability of interpersonal trust measures is sparser than in the case for institutional trust. This is because the Gallup World Poll does not include measures of interpersonal trust in the core questionnaire, while the WVS country coverage only partially overlaps with that of the ESS and the EQLS for the post-2002 period covered by the OECD Trust database. Nonetheless, for those comparisons that are possible, the picture of reliability that emerges is fairly good. Figure 2.1 shows two comparisons for generalised trust that are possible using the OECD Trust Database. The ESS and the EQLS produce very consistent results, with a correlation coefficient of 0.82. In addition, the absolute levels of generalised trust for each country are very similar across the two surveys (Panel A). The correlation between levels of generalised trust in the WVS and the Eurobarometer is a little lower at 0.75, but still relatively high (Panel B).

Although lack of data in many of the surveys included in the OECD Trust Database limits the number of comparisons that can be made between surveys, a second strategy for investigating reliability is possible using the EQLS and ESS. This involves looking at correlations between waves of the same survey. If the measure used is reliable, then correlations between successive waves should be relatively high, and the correlation should decline between waves that are more separated in time. The WVS is not suitable for this approach due both to the long period between waves and to the changes in country coverage between waves. However, both the EQLS and ESS have smaller gaps between waves (two and three years respectively), and cover a consistent panel of countries.

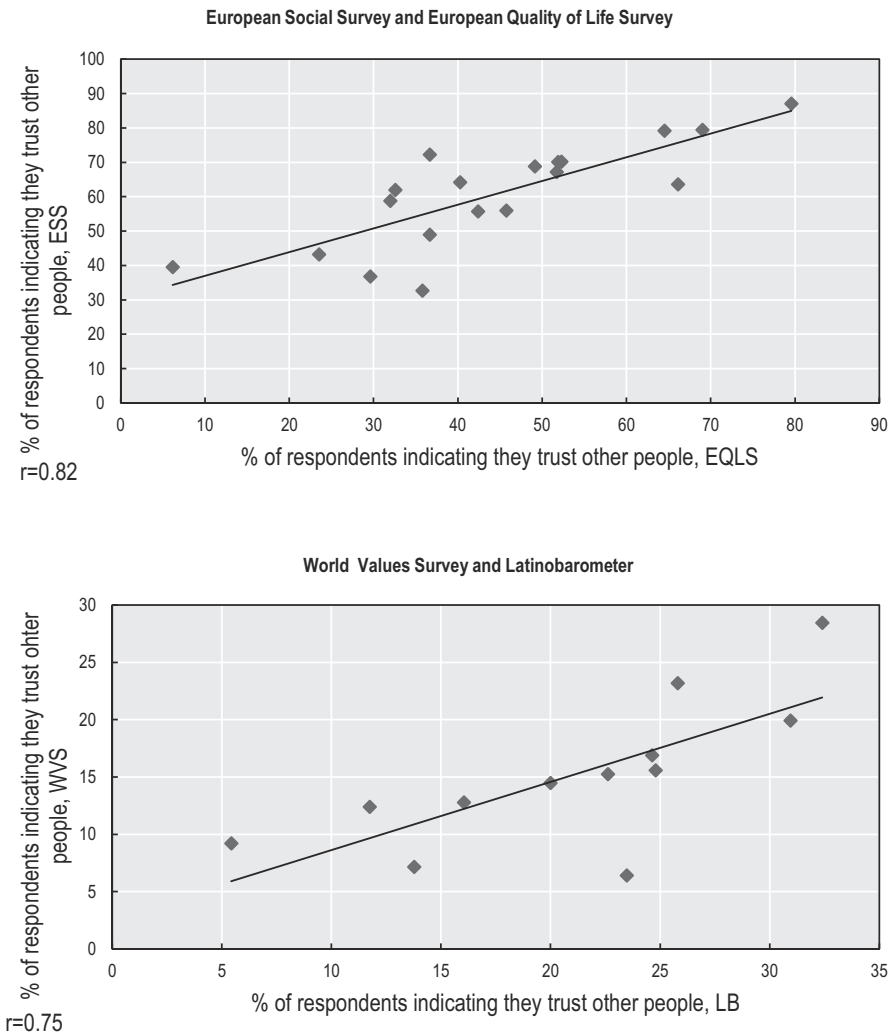
Table 2.6 below shows the correlation between waves of the ESS. It can be seen that the correlations are very high (between 0.96 and 0.98 for successive waves). The correlation falls, as would be expected, with waves that are more separated in time. Nonetheless, these correlations also remain high, and the combination of high inter-wave correlations falling with time is consistent with a high level of reliability.

Generalised trust measures are available for fewer waves of the EQLS than the ESS. Nonetheless, it is possible to look at the correlation between waves 2 and 3 of the EQLS. Figure 2.2 illustrates this and shows a high correlation of 0.93. This is a little below that found for the ESS, even allowing for the slightly longer gap between waves in the EQLS when compared to the ESS. However, this level of correlation is still high in absolute terms and is consistent with a high degree of reliability in the measure.

Face validity

Face validity addresses the degree to which a measure is intuitively plausible. This is important as it can impact on the quality and uptake of data. Poor face validity affects the quality of data because respondents may be unwilling to provide a well-considered answer

Figure 2.1. **Correlation between measures of generalised trust across surveys**



Note: Both the European Social Survey (ESS) and the European Quality of Life Survey (EQLS) responses have been recoded so that a response from 5-10 on the 0-10 response scale (ESS) and a response from 6-10 on the 1-10 response scale (EQLS) counts as indicating trust in other people.
Source: Calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933583576

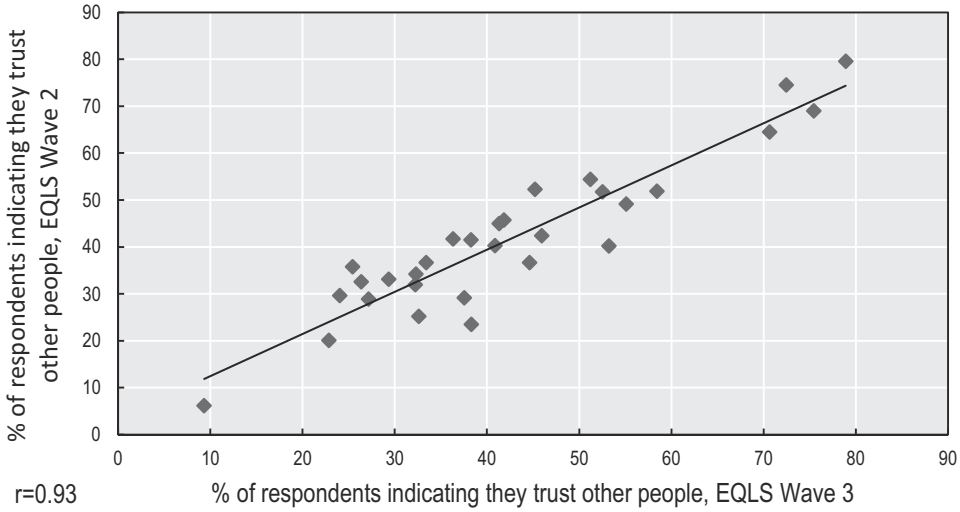
Table 2.6. **Intra-wave correlation for generalised trust, based on the European Social Survey**

Wave	1	2	3	4	5	6
1						
2	0.98					
3	0.98	0.97				
4	0.97	0.97	0.98			
5	0.94	0.94	0.98	0.96		
6	0.94	0.95	0.96	0.95	0.98	

Note: The correlation is based in the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"
Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* as contained in the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584127

Figure 2.2. **Intra-wave correlation for generalised trust based on the European Quality of Life Survey**



Note: The correlation is based on the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Responses have been recoded so that a response from 6-10 on the 1-10 response scale counts as indicating trust in other people.
Source: OECD calculations based on Eurofund (2016): European Quality of Life Survey (database) *https://discover.ukdataservice.ac.uk/catalogue/?sn=7316&type=Data%20catalogue* as contained in the OECD Trust Database.

StatLink *http://dx.doi.org/10.1787/888933583595*

to a question that seems frivolous or lacking in meaning. Similarly, data users may be unwilling to place faith in evidence that looks implausible or difficult to interpret. Hence, although face validity is a relatively "loose" and imprecise concept that is difficult to quantify, it is nonetheless of fundamental importance in establishing whether data are fit for purpose. There exists no obvious metric for assessing the face validity of a survey question, so discussion of face validity must either use qualitative or anecdotal evidence or make inferences from the behaviour of respondents. Both approaches to assessing face validity can be applied to measures of trust.

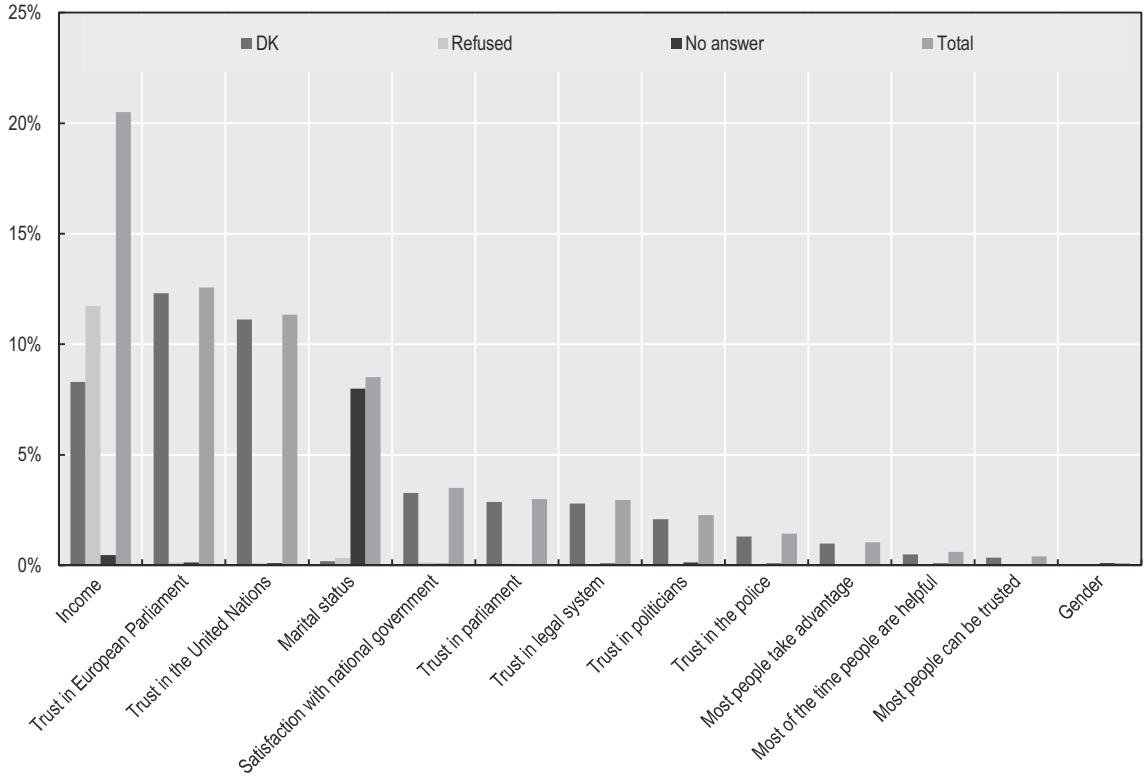
On the qualitative front, the key issue is whether the measures of trust in use seem intuitively plausible. This is a subjective judgement, and not one for which a definitive answer is possible. A number of objections to the wording of the standard trust questions have been raised over the years, notably by Hardin (2004) and Delhey et al. (2011). In both cases, it is argued that the standard Rosenberg question on generalised trust is too general to allow for a meaningful answer. Hardin also raises similar issues with respect to questions on trust in institutions, arguing that, in order for respondents to provide a meaningful answer, such questions should be specific not only about the institution considered but also in terms of what activity is expected of it.

While these issues are important, the standard questions have been considered as meaningful by the large number of authors who work with them or who have opted to use them in surveys. Knack (2001) notes that countries with high measured levels of interpersonal trust using these questions are "the Nordic nations, where citizens commonly leave unlocked bicycles and unattended strollers in public areas". While these considerations, in and of themselves, do not allow conclusive judgement on the face validity of trust measures, a more quantitative approach is to look at how difficult respondents find it to answer questions on trust. A simple way of evaluating this is to look at the item-specific

non-response rates for measures of trust compared to other widely used survey items. If respondents find the trust questions confusing or difficult to answer, then a higher item-specific non-response rate should be anticipated for these items.

Figure 2.3 shows the item-specific non-response rates for a range of different survey questions included in the European Social Survey. This includes questions on institutional trust (e.g. trust in the European Parliament, the United Nations, national parliament, the legal system and the police), questions related to generalised trust (e.g. most people take advantage, most of the time people are helpful, most people can be trusted) and a number of questions on topics unrelated to trust that are commonly asked in household surveys (e.g. income, marital status, gender). Income is, by a large margin, the most difficult topic for respondents to answer, as it requires combining information from very different income sources paid to different members of the same household; over 20% of respondents either refuse to answer or reply *don't know*. By way of contrast, only 8% of respondents did not answer questions on marital status, while gender has a response rate of close to 100%.

Figure 2.3. **Item specific non-response in the European Social Survey, 2002-14**



Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* as contained in the OECD Trust Database

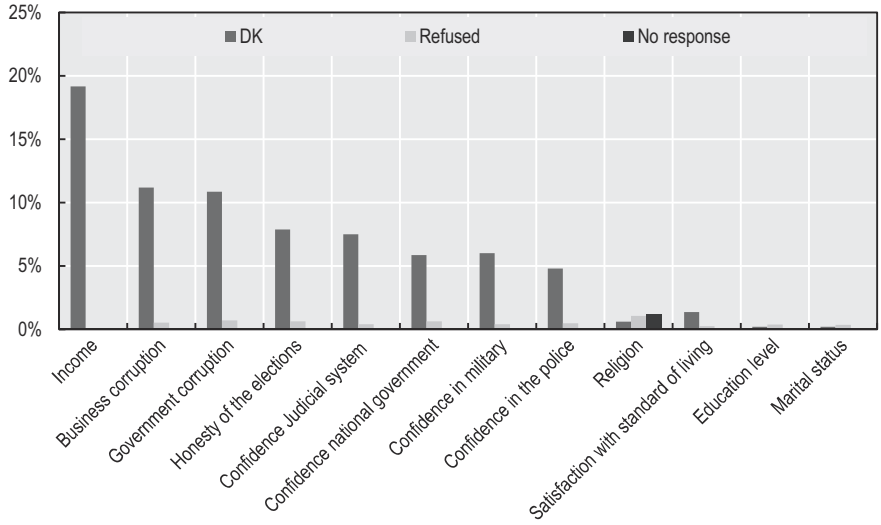
StatLink http://dx.doi.org/10.1787/888933583614

In comparison to gender and marital status, questions on institutional trust did not fare particularly well. More than 10% of respondents failed to provide an answer to questions about the European Parliament or United Nations, and even the more familiar national institutions had non-response rates of between 1% and 3%. These rates are much lower than

in the case of income, but well above those for gender. Questions on generalised trust performed better, with non-response rates of 1% or less for the three questions included here.

Figure 2.4 provides further evidence on the face validity of institutional trust by extending the analysis of non-response rates to the Gallup World Poll. This survey has a significantly larger sample than the European Social Survey, and includes a much broader cross-section of countries. The picture that emerges from the GWP analysis is very similar to that from the ESS. Income is the item with the highest item-specific non-response rate, at 19%, while marital status and education both have very low non-response rates (at around 0.50%), and religion has a non-response rate of 3%. All of the institutional trust measures in the GWP have high item-specific non-response rates, ranging from 5% for confidence in the police to 8% for confidence in the judiciary; questions on corruption and the honesty of elections have even higher item-specific non-response rates.

Figure 2.4. **Item-specific non-response rates in the Gallup World Poll, 2006-16**



Note: The response option "No response" for the question on religion was only given to respondents before 2011. Income data refers to Gallup World Poll waves 1 to 9. Missing responses were coded as "Don't know" in the case of income.
Source: OECD calculations based on the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* as contained in the OECD Trust Database.

StatLink *http://dx.doi.org/10.1787/888933583633*

The overall picture on face validity that emerges from an analysis of non-response rates is, hence, mixed. While all trust questions perform better than income – which is commonly collected in almost all household surveys – the item-specific non-response rates are higher for the bulk of the trust questions than for questions on marital status, education or gender. Even religion, which might be considered a relatively sensitive question, had a non-response rate that was less than half of most institutional trust questions (although still a substantially higher non-response rate than is the case for generalised trust in the ESS). An analysis of non-response rates does not, in itself, provide conclusive evidence on face validity. High non-response rates for measures of institutional trust could simply reflect that, although the question is reasonable, many people lack adequate information to provide a meaningful answer. The very high non-response rates for questions on less familiar organisations such as the United Nations and the European Parliament, when compared to national institutions, provide some support for this view.

Convergent validity

A measure is said to have convergent validity if it correlates well with other measures of the same construct. For example, a self-report measure of individual health status might be considered to have good convergent validity if it is correlated with the use of health services (higher usage being a proxy of poor health) and with a third-party assessment of the subject's health status. Convergent validity is easier to assess for areas where a range of plausible proxy measures of the same underlying construct exists, but for which no ideal measure exists (in which case, contingent validity would not be relevant). In the case of trust, no unambiguous direct measure is obviously available, but a number of proxies exist. This is the case for measures of interpersonal trust and particularly generalised trust, where a range of both experimental and non-experimental measures of trust can be used to test convergent validity. For institutional trust, however, the available evidence on convergent validity is sparser.

Interpersonal Trust. There are two primary sources of data on interpersonal trust that can be used to assess convergent validity. The first of these involves looking at the correlation between different survey questions on trust. Section 2.3 identified several approaches to asking questions on trust (evaluation, expectation and experience), and the correlation in responses to these different types of question provides one potentially valuable source of information. It is also possible to look at how responses from different sources vary with respect to trust. The second main source of information on convergent validity, and potentially the strongest source of evidence, lies in the results of experimental studies. There is now a large body of experimental data that can be used to validate survey questions on trust. Finally, although more limited in scope, there is a small body of evidence on the bio-physical correlates of trust that is also relevant to assess convergent validity. Almost all of this evidence focuses on generalised trust, both because data on generalised trust are more widely available than is the case for limited trust and because generalised trust is of greater relevance than limited trust for most policy and analytical purposes.

Knack (2001) provides a good overview of the validity of generalised trust measures from the perspective of convergent validity. He notes, in particular, that data from the *Reader's Digest* lost wallet experiments (Box 2.2) correlate well with survey measures of generalised trust. Despite the relatively low sample size for the lost wallet experiment, the proportion of wallets returned correlates with country values of the WVS measure of generalised trust at 0.65 ($p < 0.01$). When per capita income is controlled for, the correlation is even stronger. This finding is reinforced by Helliwell and Wang (2010), who note that the proportion of lost wallets returned in the *Reader's Digest* experiment correlates well with data from an expectations question on whether a lost wallet would be returned, which in turn correlates well with the WVS generalised trust question. This is also true for other generalised trust questions. For example, Falk et al. (2015) note that their preferred measure of trust ("I assume that people have only the best intentions") is correlated with the WVS measure of interpersonal trust at 0.53 ($p < 0.01$).

Looking at a wider range of questions, Knack (2001) notes that responses to the WVS question on generalised trust are strongly correlated with items from the same survey relating to respondents' attitudes towards taking advantage of others (e.g. cheating on taxes or not reporting damage to a parked vehicle). These correlations focus on people's assessment of their own trustworthiness rather than on whether other people can be trusted, so the measures are sufficiently different to add substantial information. Knack and Keefer (1997) note that the relationship between generalised trust and attitudes towards

taking advantage of others, like that between trust and the proportion of lost wallets returned, is stronger after controlling for per capita income. Naef and Schupp (2009) look at the relationship between people's past trusting behaviour (e.g. lending personal possessions, lending money, leaving the door unlocked) and measures of generalised trust. Using data from the German Socio-economic Panel, they find a robust relationship between generalised trust and past instances of trusting behaviour.

Another source of evidence of convergent validity is the correlation between country-average levels of generalised trust and evaluations by foreigners of how trustworthy people from different countries are. Knack reports that, using Eurobarometer data on *how much you would trust people from different countries*, there is a 0.45 correlation ($p = 0.056$) with generalised trust measured in the WVS.

Looking beyond survey data, there is now a large body of experimental evidence on the validity of measures of generalised trust. This rests largely on lab experiments using one or more variants of the Trust Game (Berg and McCabe, 1995), as described in Box 2.2. Glaeser et al. (2000) provided the first systematic use of laboratory experiments to validate survey measures of trust. Using a sample of 189 Harvard students, the authors found no significant relationship between the standard WVS measure of generalised trust[11] and trust as measured in the Trust Game (the proportion of the starting endowment sent by player 1 to player 2). While this might seem strong evidence against the validity of trust measures, Glaeser et al. also find that the survey-based measures of generalised trust are a strong and significant predictor of *trustworthy* behaviour in the Trust Game (the proportion of player 2's net endowment returned to player 1, see Box 2.2). In other words, the experimental evidence suggests that the standard generalised trust questions do collect valid data about a person's trustworthiness, but not about an individual person's willingness to trust *ex ante*. Lazzarini et al. (2004) repeat the same experiment in Brazil, while also investigating the impact of a face-to-face set-up, as opposed to an anonymous set-up for the Trust Game. Their results confirm Glaeser et al.'s findings that survey-based measures of generalised trust correlate with individual trustworthiness in the Trust Game, but not with the individual's level of trust. Cox (2004) and Capra et al. (2007) explore the relationship between survey-based measures of trust and experimental results in more detail. By including measures of other-regarding preferences – both experimental and survey-based – these authors show that trusting behaviour in the Trust Game is predicted well by the standard WVS question once altruism is controlled for. Finally, Gachter et al. (2004) report that the standard generalised trust question is associated with co-operation in the public goods game.

All these studies share one limitation: they use a small and largely unrepresentative sample of participants in the experimental games. This raises the issue of whether the results can be extended to the population as a whole. While there are currently no cross-country experimental studies involving large-scale nationally representative samples, an alternative approach is to look at the results of a systematic meta-analysis of experimental studies. Johnson and Mislin (2011, 2012) undertake a thorough meta-analysis of experimental studies, involving the Trust Game, covering 162 replications of the Berg and McCabe trust experiment, 35 countries and over 23 000 respondents. Although most of these studies are small (the average sample size is 148), they cover a wide range of countries, both developing countries (e.g. Cameroon and Uganda) and developed countries (e.g. the United States and Sweden). Contrary to earlier experimental studies, Johnson and Mislin find a significant positive correlation between the WVS measure of generalised trust and trusting behaviour in experimental games, but no relationship with trustworthiness. One

explanation for this apparent contradiction is that the authors consider the relationship between country-average levels of trust in both survey responses and experimental results, while the studies cited earlier look at individual-level correlations.

A final source of information to assess the convergent validity of generalised trust measures is provided by Fehr (2009), who discusses a series of experiments analysing the effect of oxytocin (a neuro-transmitter highly associated with pro-social behaviour in mammals). In an experimental set-up, players of the Trust Game who received a nasal spray containing oxytocin immediately before the game showed significantly higher levels of trusting behaviour than those who received a placebo spray. Fehr argues convincingly that, in this experimental design, one can effectively rule out the possibility that oxytocin affected trust via affecting players' general altruism or their risk preferences. This study hence suggests instead that the measures of trust produced by the Trust Game capture genuine trusting behaviour, and are strongly grounded in a neurophysical mechanism.

Institutional Trust. While there are more survey data available on trust in institutions than is the case for interpersonal trust – thanks to the inclusion of institutional trust measures in the Gallup World Poll, Eurobarometer and Latinobarometer – there is much less information on the convergent validity of institutional trust measures. This is particularly the case with respect to experimental analyses, as the standard Trust Game does not lend itself to the analysis of trust in institutions. Another complicating factor is that, while for interpersonal trust there is a clear consensus on using generalised trust as the preferred measure of interpersonal trust, institutional trust covers several dimensions, all of which are of interest (cf. Box 2.1).

Bouckaert and Van de Walle (2003) review the public administration literature with respect to trust and argue that measures of institutional trust do not adequately capture good governance, although their analysis is more discursive than empirical. Davidov and Coromina (2013) take a more empirical approach and use data from four waves of the European Social Survey to test measurement invariance across European countries, in both interpersonal and political trust. While they find limited support for scale invariance for measures of interpersonal trust (similar to the findings of Reeskens and Hooghe, 2008, who used metric equivalence tests on the trust item in the European Social Survey 2002 and 2004 waves), the evidence provides less support with respect to measures of political trust. In particular, their model rejects strong convergence between measures of trust in politicians and trust in parliament or the legal system. Cook and Gronke (2005) find similar ambiguity with different measures of trust in institutions, reporting correlations ranging from 0.22 to 0.36 between different questions. While Cook and Gronke view these correlations as acceptable, given that they capture slightly different concepts, they are nonetheless low when compared to the correlations between different measures of interpersonal trust. There is preliminary evidence that compares survey and experimental measures of institutional trust, relying on psychometric implicit association testing (see Box 2.4 for a description of the OECD's *Trustlab* project, or Intawan and Nicholson, 2017). The initial results, for trust in the national government, point towards a low but significant correlation between both types of measures, with implicit trust consistently scoring higher than self-reported trust.

New approaches to validating measures institutional trust. Although the existing evidence on the convergent validity of measures of institutional trust is limited, this does not mean that it is impossible to do better. In fact, there are a number of empirical strategies open

Box 2.4. **Trustlab**

A key limitation of the existing experimental literature on Trust Games is that almost all the studies draw on small, non-representative samples. Often, respondents are drawn from convenience samples such as undergraduate or graduate students in a particular course, raising concerns about the extent to which any study results would hold in a more general setting. Despite this, experimental techniques are attractive, both because they seem to predict behaviours in the real world rather well (Karlan, 2005; Benz and Meier, 2006; Algan and Cahuc, 2013) and because a different set of methodological biases is associated with them compared to those associated with the standard survey questions. This means that the two approaches are well suited to be used for validating each other. While meta-analyses such as that of Johnson and Mislin (2011) go some way to addressing these issues, and some authors (Fehr et al., 2003; Falk et al., 2016) have conducted representative surveys based on validation from small experimental samples of students, the ideal approach would involve a cross-country study involving large nationally representative samples using the same questionnaire and experimental design in each country.

Trustlab is a joint initiative between the OECD, the Paris Institute of Political Studies (Sciences Po) and a range of additional academic and governmental partners to carry out a series of comparable experimental studies of trust across OECD countries. Each study is based on a nationally representative sample of 1 000 individuals stratified by age, gender and income. In the current set-up, respondents log onto a custom-made online platform to participate in three experimental games to provide behavioural measures of social norms and values. These are: the Dictator Game, which provides information on altruism; the Trust Game, which provides information on trust and trustworthiness; and the Public Good Game, which provides information on willingness to co-operate and contribute to public goods. In all three cases, respondents are matched with other respondents for the games, which are played with approximately USD 10 in actual money at stake. In some countries, Conditional Trust Games that assess trust in specific population groups are included. After the games, respondents complete a series of implicit association tests on attitudes towards a range of institutions (namely, the government, the judicial system and the media). Implicit association tests 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 (Greenwald et al., 2002). These tests have been applied successfully to measure perceptions, stereotypes and attitudes towards commonly stigmatised social groups such as Black people, women and the elderly (Dasgupta and Asgari, 2004; Aberson et al., 2004). The final part of *Trustlab* involves respondents completing a survey questionnaire containing a battery of trust questions. These cover a range of different approaches to measuring both interpersonal and institutional trust, self-reported items on other social norms such as altruism and reciprocity, as well as a range of questions on the potential policy drivers of trust, along with basic demographic and socio-economic information. *Trustlab* has so far been run on nationally representative samples in France, Korea, the US, Germany, Italy and Slovenia, with additional countries lined up to join. The first evidence is expected to become available at the beginning of 2018.

for investigating the convergent validity of these measures that might yield useful information. Most of these involve analysing existing data on institutional trust in the context of actual behaviour and evaluating whether the level of trust revealed by people's behaviour is consistent with survey measures of institutional trust. The fact that these have not to date

been the subject of any in-depth analysis represents a significant misallocation of scholarly resources. In particular, although there is a large political science literature analysing aspects of institutional trust, the systematic testing of the validity of institutional trust measures against actual behaviour is a major research gap. Although it is beyond the scope of these Guidelines to provide a detailed breakdown of all possible research strategies for testing the validity of measures of institutional trust, some proposals for research are set out below in the interests of ensuring that scholarly effort is directed towards issues of high importance.

Trust in the police, and in the legal system more broadly, are two areas where it should be possible to obtain information on convergent validity. Measures of trust both in the police and in the legal system are available from a number of surveys with good cross-country coverage (e.g. the ESS, EQLS, WVS): this provides the survey data for analysis. The behavioural data against which to test the validity of survey measures could come from victimisation surveys and police data. In particular, high trust in the police and the legal system should be reflected in a higher proportion of incidents of criminal victimisation being reported to the police. Victimisation surveys provide information on the base rate of criminal victimisation, while police statistics record the rate at which incidents of victimisation are reported to the police. It should therefore be possible to examine whether the ratio of incidents of criminal victimisation reported to the police as a proportion of total incidents of victimisation is closely associated with the values of trust in police across countries.

Other similar analyses are also possible to provide information on other aspects of institutional trust. For example, the per capita consumption of bottled water could be used as a proxy measure of trust in the institutions associated with the water supply and tested across different jurisdictions within a country. Clearly in such a case, controlling for the income of different areas would be essential, but this is not a major challenge. Alternatively, variance across a country in the proportion of the population with private medical insurance might be used as a proxy indicator of trust in public medical services. Other similar proxies could be identified for areas such as education or disaster preparedness. While analyses of this sort do not provide much information on overall trust in public institutions (possibly reflecting inherent difficulties in the concept), they would provide a much stronger grounding for measures of trust in specific institutions. For this reason, research into the convergent validity of institutional trust using behavioural data should be a major research priority with respect to institutional trust.

Construct validity

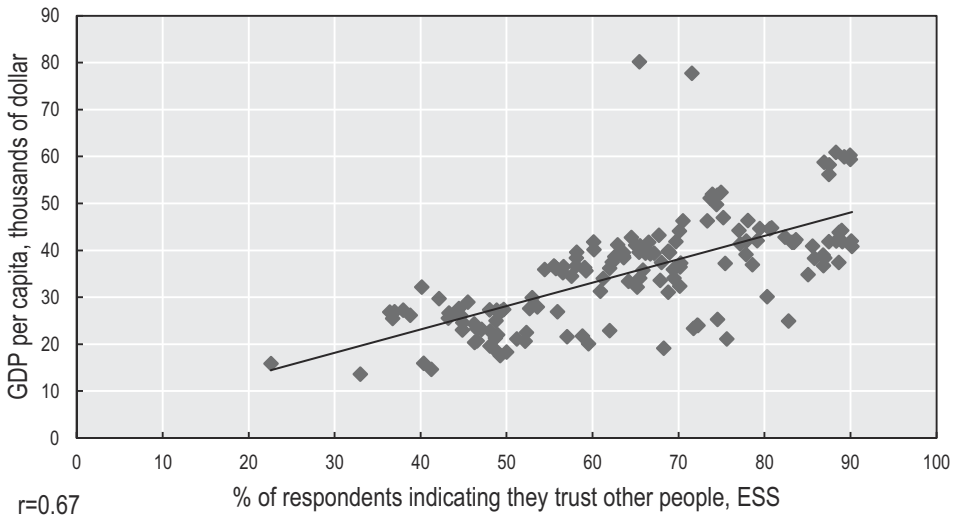
In addition to evidence that measures of trust drawn from different sources provide the same picture of the variation in trust across countries, it is also useful to look at whether trust measures behave in a way that is consistent with expectations. For example, if trust in strangers (i.e. the standard WVS question on generalised trust) is an important part of social capital (as argued by Scrivens and Smith, 2013), then one would expect to see higher levels of it associated with better economic and social outcomes. Similarly, higher levels of institutional trust should be associated with greater willingness on the part of people to co-operate with government agencies in the pursuit of common goals, or with higher measures of government trustworthiness.

Generally speaking, measures of both interpersonal and institutional trust are well supported in terms of construct validity. While only a few papers have analysed trust from the perspective of convergent validity (particularly for measures of institutional trust), a large literature has looked at the determinants and correlates of trust – both interpersonal and

institutional. This reflects the high degree of interest in measures of trust within economics, political science, sociology and public management. Indeed, the fact that trust measures perform so well in terms of construct validity helps to explain why there has not been a stronger emphasis on more formally testing the other types of validity of such measures.

Interpersonal Trust. For interpersonal trust, within the limits of existing unofficial data, convergent validity is well supported, in particular for generalised trust. Countries with high levels of generalised trust tend to have higher incomes (Fukuyama, 1995; Whitely, 1997; Knack and Keefer, 1997; La Porta et al., 1997; Algan and Cahuc, 2013), which is consistent with the view that generalised trust is a core element of social capital and is important for lowering transaction costs between people who do not know each other well. Figure 2.5 below illustrates the relationship between income (GDP per capita) and generalised trust in the ESS between 2002 and 2014, showing a clear correlation between the two measures (r of 0.67) for European countries.

Figure 2.5. **Generalised trust and GDP per capita in European countries, 2002-14**



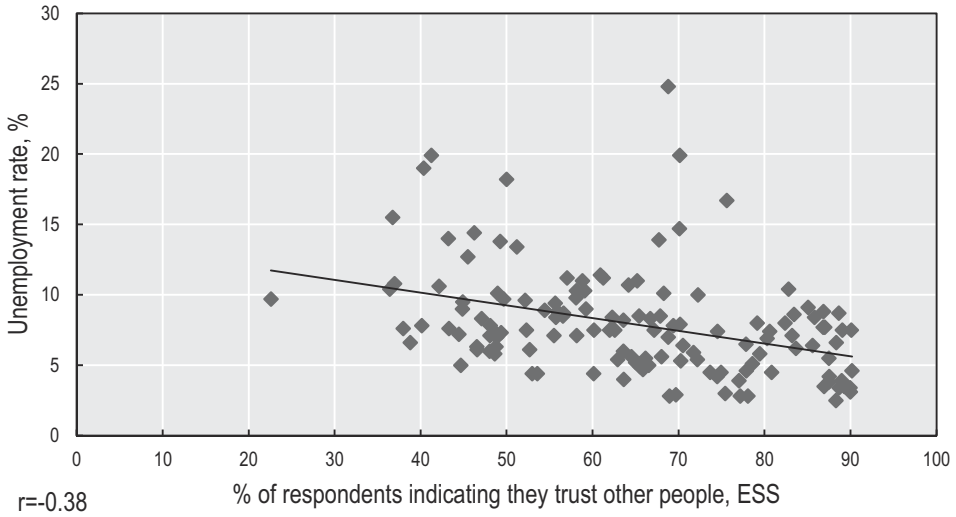
Note: The correlation is based on the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Responses have been recoded so that a response from 5-10 on the 0-10 response scale counts as indicating trust in other people.
Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* and OECD Stat (database) *http://stats.oecd.org/index.aspx?DataSetCode=PDB_LV* as contained in the OECD Trust Database.

StatLink *http://dx.doi.org/10.1787/888933583652*

In addition to levels of income, trust is also correlated with the short-term performance of the economy. Figure 2.6 shows the correlation between generalised trust in the ESS and unemployment rates in European countries between 2002 and 2014. Although the relationship is weaker than for income, generalised trust is significantly negatively correlated with unemployment rates across European countries.

There is also evidence that generalised trust correlates positively with better health outcomes for individuals (Kawachi and Berkman, 2000; Boreham et al., 2004; Ginn and Arber, 2004). Figure 2.7 shows the relationship between generalised trust and life expectancy using the same dataset as Figures 2.5 and 2.6. Although the relationship is weaker than is the case for income, life expectancy is still strongly correlated with generalised trust (r of 0.44).

Figure 2.6. **Generalised trust and unemployment in European countries, 2002-14**



Note: The correlation is based on the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Responses have been recoded so that a response from 5-10 on the 0-10 response scale counts as indicating trust in other people.
Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* and OECD Stat (database) *http://stats.oecd.org/index.aspx?queryid=36324* as contained in the OECD Trust Database.
StatLink *http://dx.doi.org/10.1787/888933583671*

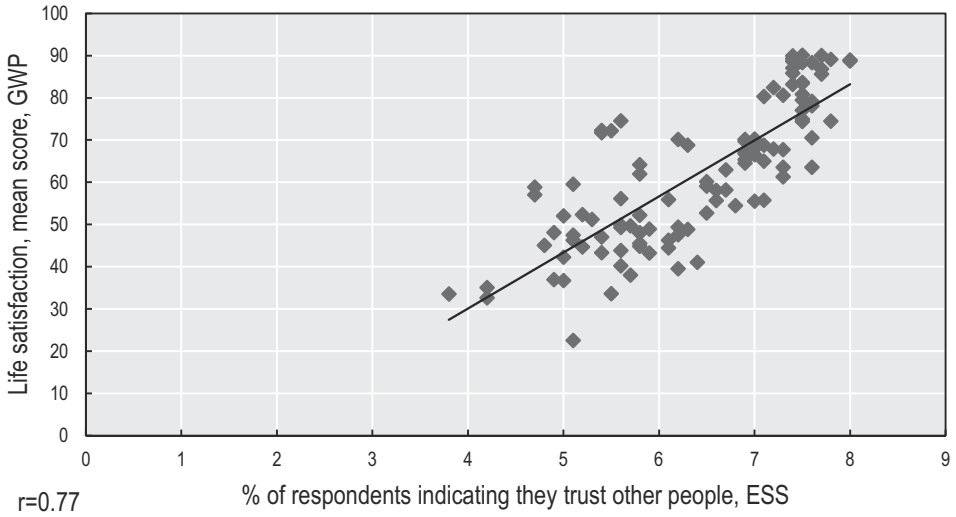
Figure 2.7. **Generalised trust and life expectancy in European countries, 2002-14**

Note: The correlation is based on the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Responses have been recoded so that a response from 5-10 on the 0-10 response scale counts as indicating trust in other people.
Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* and OECD Stat (database) *https://stats.oecd.org/index.aspx?queryid=24879* as contained in the OECD Trust Database.
StatLink *http://dx.doi.org/10.1787/888933583690*

There is also evidence of a robust relationship between generalised trust and other important social outcomes. As with incomes, this is consistent with the social capital literature, suggesting that generalised trust is important in allowing strangers to co-ordinate actions to achieve a wide variety of different goals. Sampson (2012) finds a negative

correlation between generalised trust and levels of crime, while there is a large literature on the relationship between generalised trust and measures of subjective well-being (e.g. Algan and Cahuc, 2013; Helliwell and Wang, 2010). Figure 2.8 illustrates the correlation between generalised trust and life satisfaction at the country level, showing that variation in generalised trust co-varies strongly with variation in life satisfaction across European countries (r of 0.77). Boarini et al. (2012) take this analysis further and show that average levels of interpersonal trust at the country level are strongly correlated with the life satisfaction of the inhabitants of these countries, independently of the individual's own trust, and after controlling for demographic and economic variables.

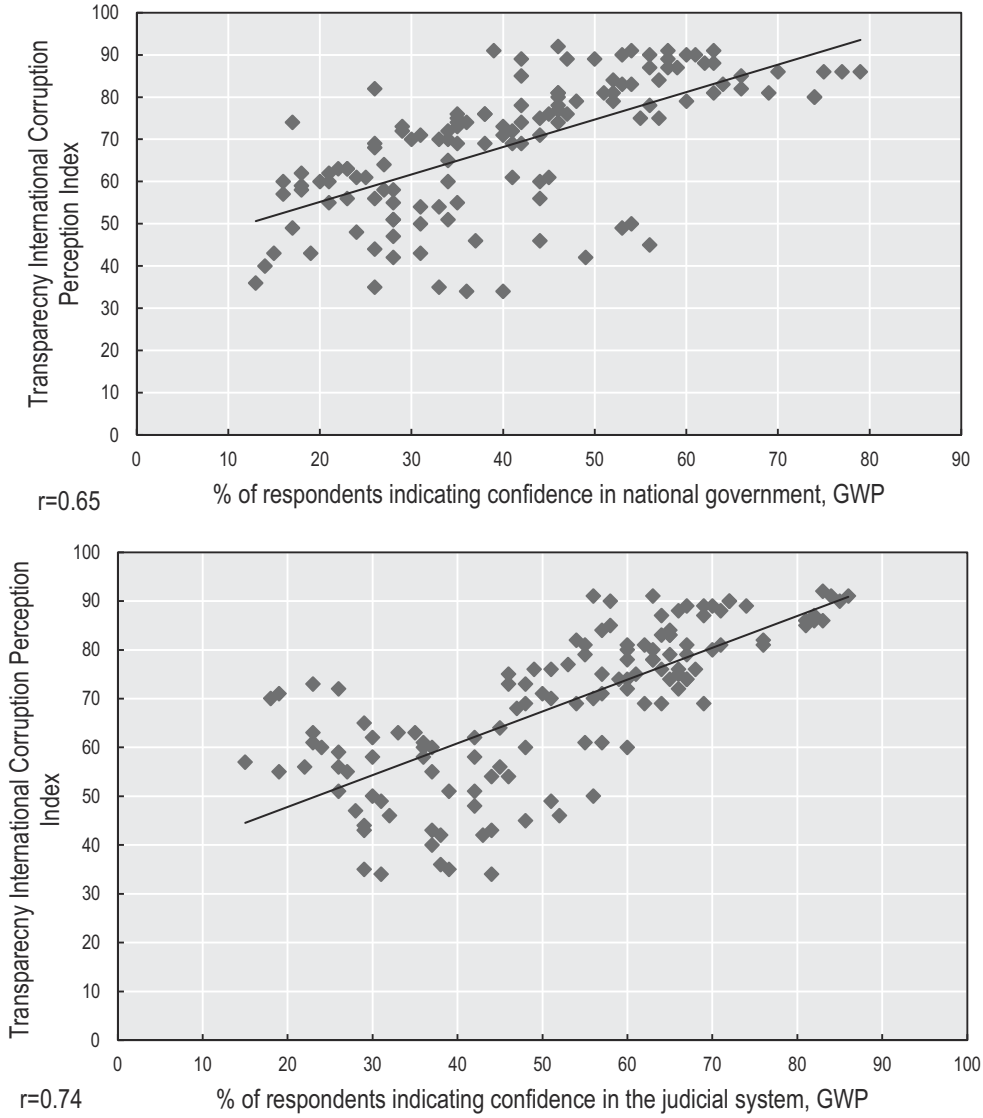
Figure 2.8. **Generalised trust and life satisfaction in European countries, 2002-14**



Note: The correlation is based on the standard generalised trust question: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" Responses have been recoded so that a response from 5-10 on the 0-10 response scale counts as indicating trust in other people. Life satisfaction scores refer to a country's average score on a response scale from 0-10 for the question: "Overall, how satisfied are you with life as a whole these days?" where zero means "not at all satisfied" and ten means "completely satisfied".
Source: OECD calculations based on the European Social Survey (database) *www.europeansocialsurvey.org/data/* and the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* as contained in the OECD Trust Database.
StatLink *http://dx.doi.org/10.1787/888933583709*

Institutional Trust. There is good evidence of a positive relationship between institutional trust and citizen support for government policy. In a cross-country analysis, Zhao and Kim (2011) highlight a positive correlation between institutional trust and levels of the Foreign Direct Investment that a country receives. Murphy (2004) and Murphy et al. (2009) find a significant positive relationship between trust in regulators and voluntary compliance in the area regulated; similarly, Daude et al. (2012) have documented a strong link between institutional trust and willingness to pay taxes. Furthermore, Knack and Keefer (1997) analysed responses to the WVS across about 30 countries, finding a positive correlation between measures of citizens' confidence in government and objective indicators of bureaucratic efficiency. Moreover, the German Satisfaction with Government Services Survey (Zufriendenheitsbefragung, 2015) points to a positive correlation between satisfaction with services, on one side, and measures of competences of a specific agency and trust in that office, on the other. There is also a robust cross-country correlation between trust in institutions and perceptions of corruption (OECD, 2013b). Figure 2.9 illustrates this relationship

Figure 2.9. **Institutional trust and perceptions of corruption in OECD countries, 2006-15**



Note: The Gallup World Poll asks the following two questions: "In this country, do you have confidence in each of the following, or not? How about national government? How about judicial systems and courts?" Transparency International's Corruption Perception Index scores countries from 0-100. The higher the score, the more transparent/less corrupt a country is.

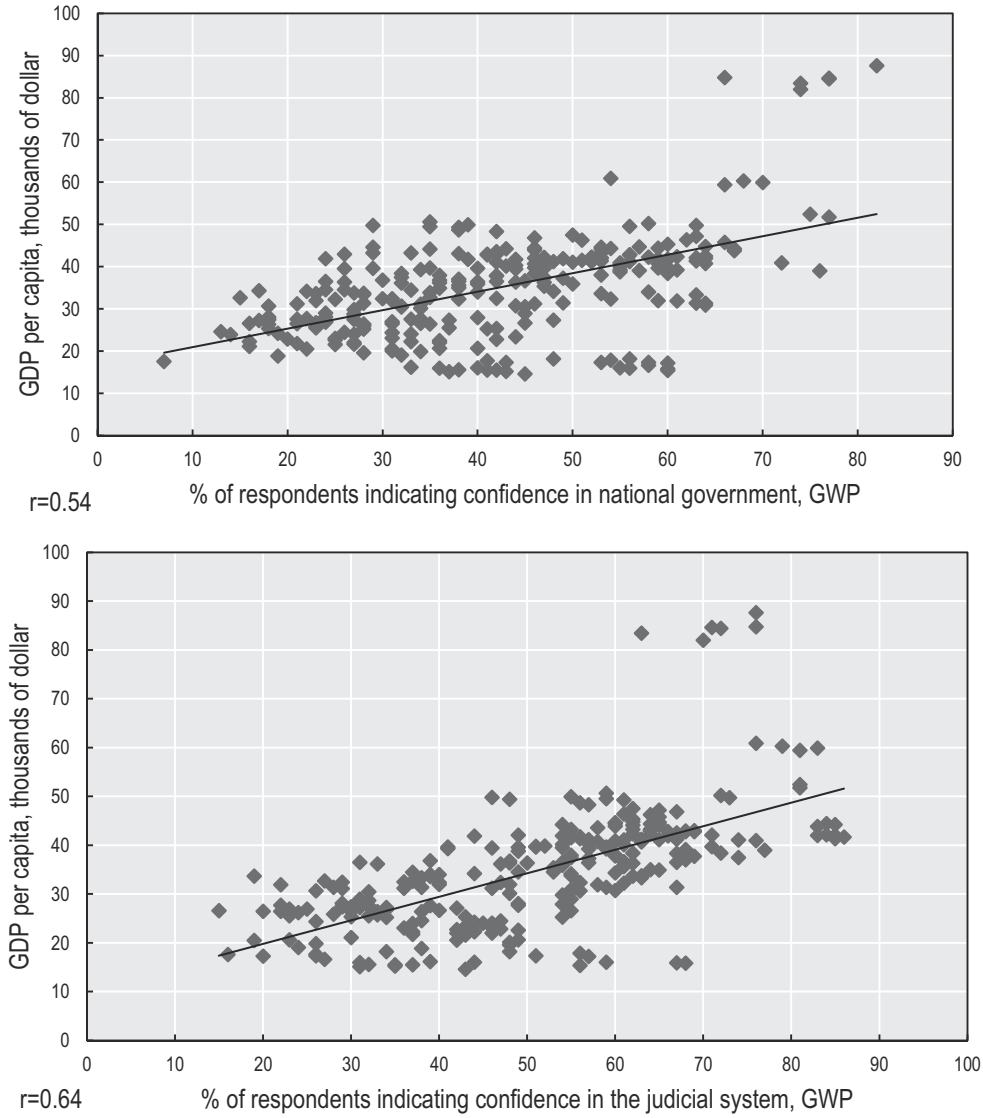
Source: OECD calculations based on the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* and Transparency International (2016): Corruption Perception Index (database) *www.transparency.org/news/feature/corruption_perceptions_index_2016* as contained in the OECD Trust Database.

StatLink *http://dx.doi.org/10.1787/888933583728*

for OECD countries, showing that perceived levels of corruption are correlated with both trust in government (r = 0.65) and trust in the judicial system (r = 0.74).

As with interpersonal trust, there is also evidence that institutional trust is linked to economic outcomes. Figure 2.10 shows the relationship between trust in two institutions – government and the judiciary – and GDP per capita. In both cases the correlation is strongly positive, slightly more so in the case of trust in the judiciary. This makes intuitive sense, as many of the key levers through which institutions affect economic outcomes, such as

Figure 2.10. **Institutional trust and GDP per capita in OECD countries, 2006-15**

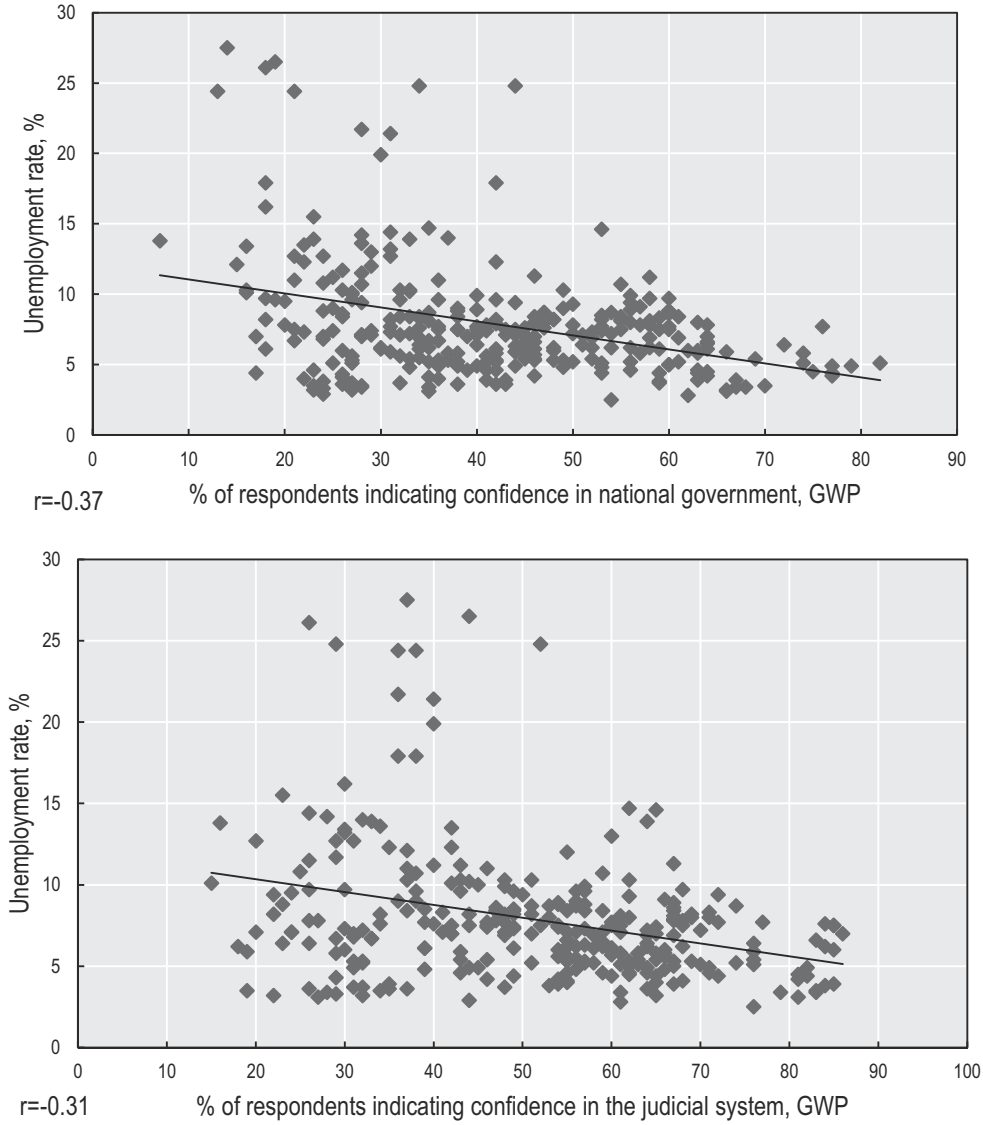


Note: The Gallup World Poll asks the following two questions: "In this country, do you have confidence in each of the following, or not? How about national government? How about judicial systems and courts?"
Source: OECD calculations based on the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* and OECD Stat (database) *http://stats.oecd.org/index.aspx?DataSetCode=PDB_LV* as contained in the OECD Trust Database
StatLink *http://dx.doi.org/10.1787/888933583747*

contract enforcement or regulation of the market place, have a more direct link to the judicial system than to the government more generally.

Roth (2015) finds a strong negative correlation between interpersonal trust and unemployment for 12 countries in the euro area from 1999 to 2014. Figure 2.11 confirms this relationship, using a different set of data sources, and also provides information on the relative strength of the relationship for different forms of institutional trust. While there is a relationship between trust in the judicial system and the unemployment rate (r =- 0.31), the relationship is stronger for trust in the government (r = -0.37). This is consistent with a view that responses to questions on trust in institutions reflect the respondent's trust both

Figure 2.11. **Institutional trust and unemployment rates in OECD countries, 2006-15**

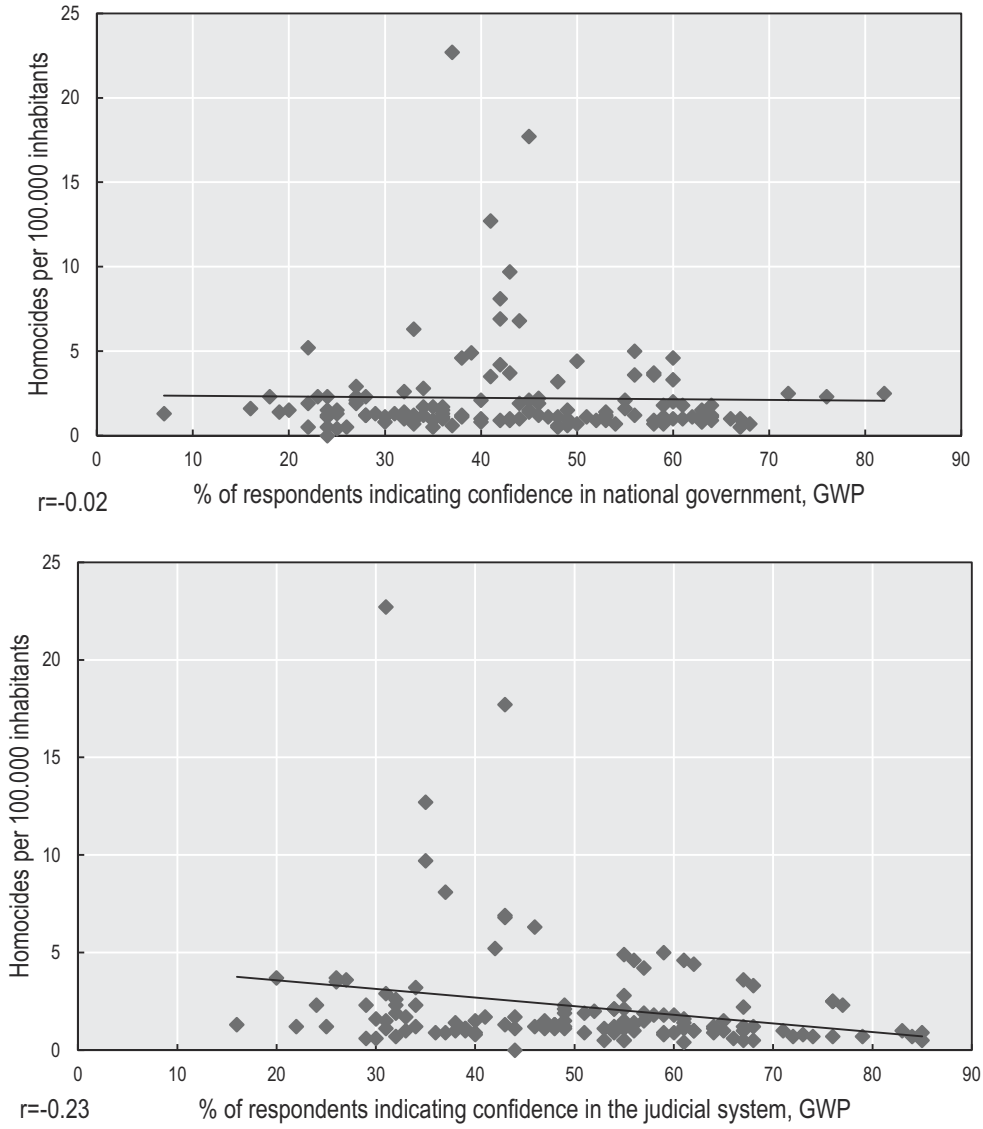


Note: The Gallup World Poll asks the following two questions: "In this country, do you have confidence in each of the following, or not? How about national government? How about judicial systems and courts?"
Source: OECD calculations based on the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* and OECD Stat (database) *https://stats.oecd.org/index.aspx?queryid=36324* as contained in the OECD Trust Database.
StatLink *http://dx.doi.org/10.1787/888933583766*

in institutions generally as well as in the specific institution that is the focus of the question. In this case, the correlation is higher for government than for the judicial system, reflecting the judicial system's weaker relevance to employment policy.

One area where one would expect to see a much stronger relationship with trust in the judicial system than with trust in the government more generally is crime. If measures of institutional trust are valid, then variation in the crime rate should be more strongly linked to trust in the judicial system than to trust in government as a whole. This is exactly what is shown in Figure 2.12. Using homicide rates per 100 000 as a proxy for the prevalence of

Figure 2.12. **Institutional trust and homicide rates in OECD countries, 2006-15**



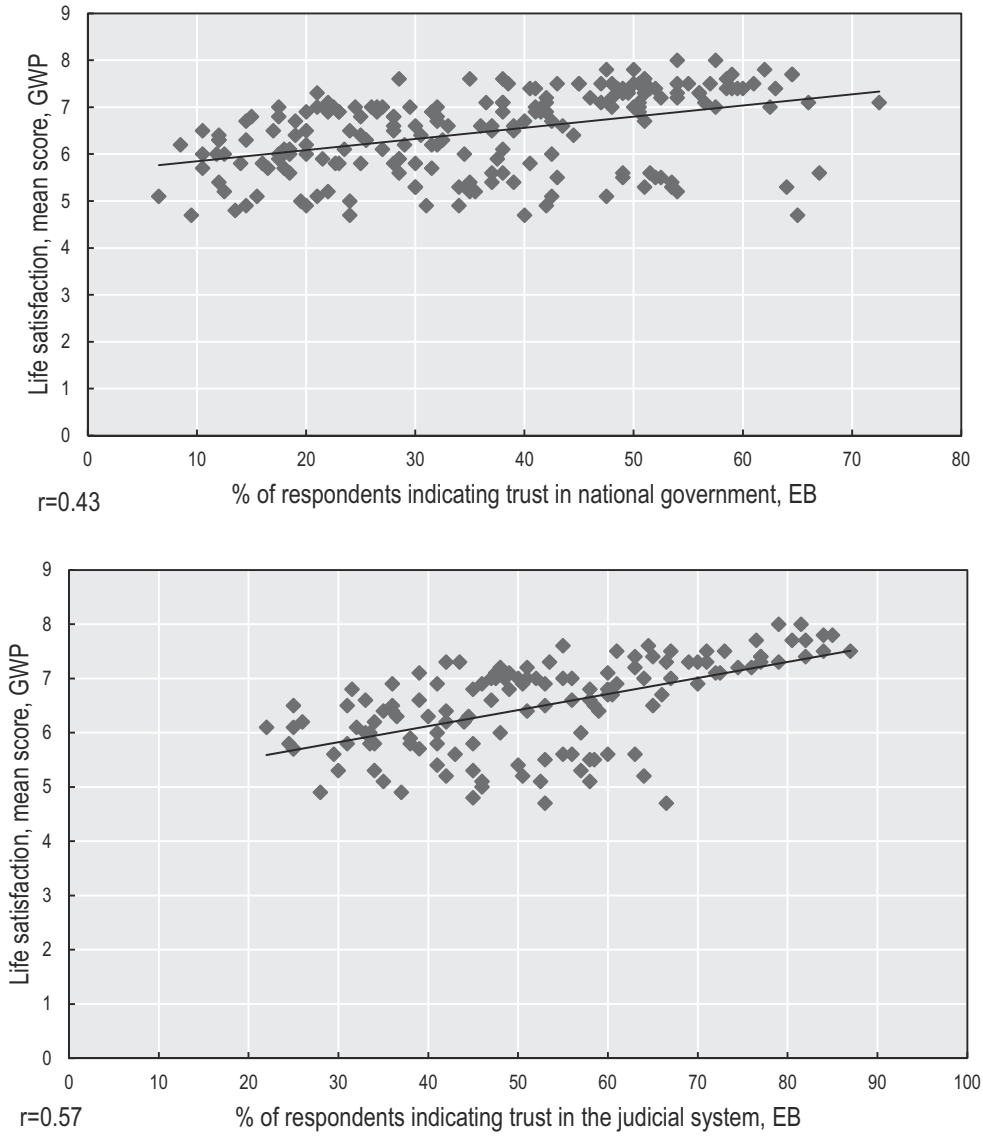
Note: The Gallup World Poll asks the following two questions: "In this country, do you have confidence in each of the following, or not? How about national government? How about judicial systems and courts?"
Source: OECD calculations based on the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* and UNDOC Homicide Statistics (database) *www.unodc.org/gsh/en/data.html* as contained in the OECD Trust Database.
StatLink *http://dx.doi.org/10.1787/888933583785*

crime, the figure shows that there is essentially no correlation between crime rates and trust in government, but a significant negative correlation with trust in the judicial system (r = -0.23).[12]

A final check on the convergent validity of measures of institutional trust is to look at the relationship between these measures and overall life satisfaction. A positive correlation between trust in institutions and life satisfaction could be expected both because trusted institutions function better and are therefore more likely to be associated with other outcomes that are important to people's life satisfaction and because trustworthy government is of direct importance for how people value their lives (Frey and

Stutzer, 2005, 2006). Figure 2.13 shows a strong positive relationship between overall life satisfaction and both trust in government and trust in the judicial system. Although this relationship is a little stronger for trust in the judicial system (r = 0.57) than for trust in government (r = 0.43), it is lower for both than is the case for the correlation between generalised trust and life satisfaction, illustrated in Figure 2.8 (r = 0.77).

Figure 2.13. **Institutional trust and life satisfaction in OECD countries, 2006-15**



Note: The Eurobarometer asks the following two questions: "I would like to ask you a question about how much trust you have in certain institutions. For each of the following institutions, please tell me if you tend to trust it or tend not to trust it? National government. The legal system. "Life satisfaction scores refer to a country's average score on a response scale from 0-10 for the question": Overall, how satisfied are you with life as a whole these days?" where zero means "not at all satisfied" and ten means "completely satisfied".

Source: OECD calculations based on the European Commission: Eurobarometer (database) *http://ec.europa.eu/commfrontoffice/publicopinion/index.cfm#p=1&instruments=STANDARD* and the Gallup World Poll (database) *www.gallup.com/services/170945/world-poll.aspx* as contained in the OECD Trust Database.

StatLink *http://dx.doi.org/10.1787/888933583804*

Key messages

Validity of measures of interpersonal trust. In general, evidence for the validity of measures of generalised trust is strong, while there is relatively little evidence focusing specifically on the validity of measures of limited trust. Measures of generalised trust perform well in terms of face validity, construct validity and convergent validity. This holds whether measures are assessed at the cross-country level or at the level of individual responses. While there is some question over the test-retest reliability of some measures of generalised trust at the individual level, country-level results are highly reliable across different data sources and over time. Although the evidence base on the validity of generalised trust measures is less developed than would be ideal, it is sufficient to conclude that such measures are fit for purpose. In particular, measures of generalised trust can provide meaningful information on population well-being and social capital and on the drivers of other social and economic outcomes. With respect to limited trust, the lack of evidence is not perhaps quite such a significant problem as it may seem at first sight. While there is little direct evidence on limited trust, most limited trust questions are conceptually very similar to generalised trust questions, suggesting that more detailed analysis of validity might yield similar positive results.

Validity of measures of institutional trust. The evidence on the validity of measures of institutional trust is more mixed than is the case of interpersonal trust. While institutional trust measures generally perform well in terms of construct validity, the picture is much less clear with respect to face and convergent validity. The available evidence on non-response rates suggests that while still performing better than measures of income (which are themselves regularly collected as part of official statistics), measures of institutional trust have higher item-specific non-response rates than measures of interpersonal trust or typical demographic questions. Non-response rates are higher for institutions with which people have little day-to-day contact such as the United Nations or the European Union, indicating that a likely cause of low response rates is lack of knowledge on the part of respondents. In the case of convergent validity, the issue is more a complete lack of evidence rather than any specific evidence pointing to a lack of validity on the part of measures of institutional trust.

The limits of validity. Most of the discussion of validity has been couched in terms of evidence in favour of the validity of trust measures. To a large degree, this reflects the balance of the academic research on the subject, as the general picture is largely positive. However, some findings do suggest significant limitations with respect to the validity of measures of trust. One potential source of concern is the finding reported by Glaeser et al. (2000) and Lazzarini et al. (2004) that survey measures of generalised trust do not correlate with the standard measure of trust in the Trust Game, but only with trustworthiness. As discussed above, however, this result is overturned when other-regarding preferences (i.e. preferences over another individual's material payoffs, in addition to one's own) are controlled for (Cox, 2004; Capra et al., 2007) or when looking at cross-country averages as opposed to individual results (Johnson and Mislin, 2012).

A more significant issue raised by Delhey, Newton and Welzel (2011) is that there are significant cross-country differences in terms of how people interpret the Rosenberg generalised trust question. In particular, Delhey et al. find that, in the majority of countries,

the "most people" element of the Rosenberg generalised trust question refers to people whom the respondent does not know, while in a significant minority of countries respondents frame their answer in terms of family and people whom they know well; this is the case in particular of poorer countries and those with a Confucian culture. While the significance of the Delhey et al. findings is disputed (Uslaner 2002; Algan and Cahuc, 2013), a non-random cross-country pattern in how the question is interpreted is a significant cause for concern; not only does a systematic difference in how the generalised trust question is interpreted increase measurement errors, it would also imply a systematic bias in results that would make cross-country or cross-group comparisons difficult to interpret.

While, compared with measures of interpersonal trust, institutional trust does not face the issues associated with trust in strangers vs. trust in those you know, Schneider (2016) raises another important problem. In an analysis of institutional trust in Eastern European and former Soviet bloc countries, Schneider notes that the error term associated with measures of institutional trust is correlated across countries with what it is attempting to measure; in other terms, there is evidence that reported trust in institutions is "U-shaped" with respect to the trustworthiness of institutions. This implies that, initially, as institutional quality falls, people will report lower levels of trust in institutions; however, if government institutions become more corrupt, respondents become unwilling to report a lack of trust in government, presumably reflecting a lack of trust in the institution collecting the information. In contrast to the issue raised with generalised trust by Delhey et al., where the actual impact on the performance of the measure appears to be limited, the problems raised by Schneider with respect to institutional trust are significant.

Over and above the three issues already discussed, there is one more significant point that needs to be stressed, i.e. the lack of evidence with respect to many of the key points relating to the validity of trust measures. With respect to generalised trust, the evidence base is generally fairly strong, and gaps are limited to a relatively narrow range of issues. In particular, more evidence on the test-retest reliability of interpersonal trust measures would be valuable, as would a more systematic examination of the non-experimental aspects of convergent validity. While Knack (2001) and Naef and Schupp (2009) provide a useful overview of the available evidence, it would be valuable to see a more systematic quantitative investigation of convergent validity issues featuring data from surveys explicitly designed to test questions related to validity. To some degree, this is already happening (Naef and Schupp, 2009; Falk et al., 2015; *Trustlab*), and there is cause to be relatively optimistic here.

In the case of institutional trust, the gap in the available evidence on validity is more severe. There is very little evidence on the convergent validity of institutional trust measures. Given that the evidence from face validity and construct validity is ambiguous, this is a major limitation. While the balance of evidence that measures of institutional trust *do not* work is not strong, studies that could provide decisive support for the validity of institutional trust measures are still lacking. More importantly, there is little to suggest that improvements in this area are currently a major area of focus. Most of the recent developments in experimental measures focus on interpersonal trust rather than institutional trust, and those experimental studies that do exist generally have limited wider applicability (e.g. Penczynski and Santana, 2016). Until this gap is filled, a lack of evidence remains a barrier to higher confidence in the validity of measures of institutional trust.

2.5. Conclusion

Key points made in this chapter are as follows:

- Interest in levels of trust in society has never been higher, and this creates a strong incentive to find ways to measure trust: both trust in other people and trust in institutions are key ingredients for societies to function. From the perspective of national statistical agencies, the key question is whether such interest should translate into a systematic effort to measure trust. Relevance and accuracy are the two key criteria for national statistical agencies to inform this decision, and it is important that any proposed official statistical measures meet both criteria. Relevance is important because it reflects whether the data collected are important to users and whether it can contribute to informing policy makers and the public more generally, while accuracy is important because data can usefully contribute to decision making only if it provides a true picture of the issues it is intended to help assess.
- The relevance of measures of trust is not in doubt. Measures of interpersonal trust – particularly generalised trust – are of fundamental importance to assessing the well-being of societies, to measuring social capital, and to understanding the drivers of other social and economic outcomes. This is reflected both in the large and growing literature on the drivers and consequences of generalised trust and in the wide range of national and international initiatives that include generalised trust as an outcome of interest, ranging from national initiatives to measure well-being such as those of the United Kingdom and Israel through to the UN Sustainable Development Goals. Similarly, there is a high level of interest in institutional trust, which is reflected both in measurement initiatives and in a large and varied academic literature. In addition to being important for measuring well-being, understanding institutional trust is essential to understanding government effectiveness and the functioning of democratic systems of government.
- The accuracy of trust measures is less clear. In general, the evidence for the validity of measures of generalised trust is strong, while there is relatively little evidence focusing on the validity of limited trust measures. Measures of generalised trust perform well in terms of face validity, construct validity and convergent validity. This holds whether the measures are assessed at the cross-country level or at the level of individual responses. While there is some question over the test-retest reliability of some measures of generalised trust at the individual level, country-level results are highly reliable across different data sources and over time. As with all intangible concepts, measuring generalised trust raises a number of issues about respondents' interpretation of the question in front of them and their subjective judgement, but these issues are not intractable. There is good evidence that, despite these issues, questions on generalised trust produce valid data, and there is extensive research that is providing new insights into the remaining measurement issues. In fact, it is likely that more is understood about the validity of measures of generalised trust than is the case for many more traditional elements of official statistics. From this perspective, measures of generalised trust can be considered as fit for purpose and should be measured in official surveys where relevant.
- The picture for measures of institutional trust is more mixed than is the case for measures of interpersonal trust. While institutional trust measures generally perform relatively well in terms of construct validity, the situation is less clear with respect to face validity and convergent validity. There is thus good reason to believe that such measures might be biased in some circumstances, and for a number of key aspects of validity there is simply

no evidence one way or the other. For this reason, despite the relevance of measures of institutional trust, it is less easy to state that they are fit for purpose within the context of official statistics. This does not mean that such measures have no place in the official statistical system, however. Rather, it suggests that such measures should be considered more experimental and should be implemented in contexts where their experimental status is clear; this is particularly important for national statistical offices. On the other hand, the relevance of such measures suggests that they should be a high priority for further research, both in the academic community and within national statistical agencies. Many of the key methodological questions regarding the accuracy of measures of institutional trust will require the sample size and response rates that only national statistical offices are able to deliver. Progress is already being made on some of these issues, which will be the focus of Chapter 3.

Notes

1. A search of the Econlit database for the terms trust or trustworthiness between 2000 and 2016 returns 5 242 records. A similar search on Google Scholar for trustworthiness returns 89 200 results over the same period. These findings illustrate both the size of the literature on trust and the degree to which looking at only a single discipline – economics – will miss much of the available evidence and understanding. Not surprisingly, a literature this diverse produces many definitions of trust.
2. Hardin (2004) does not consider the possibility of an institution being the trusting party rather than the party being trusted, but his framing could logically be extended in this way.
3. Evidence from the German Satisfaction with Government Services Survey (Zufriedenheitsbefragung 2015), which found large differences in trust between individual public agencies, supports this argument.
4. Here, trusting behaviour refers to the person engaging in trust by taking a risk on the expectation of another person's behaviour, while trustworthy behaviour consists of responding in a way consistent with expectations.
5. A range of other interactive games have been used to measure preferences and social norms other than trust; these include altruism, through the Dictator Game; co-operation and pro-social behaviour, through the public goods game; reciprocity, through the ultimatum game; and risk aversion, through lotteries. For an overview of games used in experimental economics, see Smith (2008).
6. In psychology, for instance, a survey by Arnett (2008) found that 96% of subjects in studies published in top journals were from WEIRD (Western, Educated, Industrialized, Rich and Democratic) backgrounds. Researchers – often implicitly – assume that either there is little variation in experimental results across populations, or that these WEIRD standard subjects are as representative of the human species as any other population. This is not the case, as Heinrich et al. (2010) conclude in their review of studies conducted across the behavioural sciences that WEIRD subjects are among the *least* representative populations one could find for generalising about humans, and that there is substantial variability of results across countries.
7. See also Benz and Meier (2006) for a discussion of the altruistic behaviour of Swiss students in lab experiments as compared to in the real world.
8. *Trust in different public institutions* has been suggested as a possible indicator that could be used in the future for the monitoring of Target 16.6 ("Develop effective, transparent and accountable institutions at all levels") of the UN 2030 Agenda (*https://unstats.un.org/unsd/statcom/48th-session/documents/2017-2-IAEG-SDGs-E.pdf*).
9. Measures of affect can be thought of as measures of particular feelings or emotional states such as *joy*, *contentment*, *anger*, or *worry*, and are often collected with reference to a particular point in time. In Boarini et al. (2012), affect balance is calculated as the number of positive emotions reported by a respondent on the previous day less the number of negative emotions, averaged across all respondents.
10. 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.

11. Using an index derived from the different questions on interpersonal trust – the Rosenberg question, whether people would take advantage of you if they got the chance, and would you say that most of the time people try to be helpful – Glaeser et al. (2000) find an insignificant negative relationship between survey-based measures of trust and experimental trust.

12. Although this correlation is low in absolute terms, this may reflect the limits of the homicide rate as a proxy for crime more generally: homicide is both a relatively rare crime (with rates that can vary a lot from year to year with only small changes in the actual number of homicides) and not necessarily a strong proxy for non-violent crime.

References

Aberson, C.L., C. Shoemaker and C. Tomolillo (2004), "Implicit bias and contact: The role of interethnic friendships", *The Journal of Social Psychology*, Vol. 144, No. 3.

Algan, Y. and P. Cahuc (2013), "Trust, growth and well-being: New evidence and policy implications", *IZA Discussion Paper,* No. 7464.

Algan, Y. and P. Cahuc (2010), "Inherited trust and growth", *The American Economic Review,* Vol. 100, No. 5, pp. 2060-2092.

Alkire, S. (2002), "Dimensions of Human Development", *World Development,* Vol. 30, No. 2, pp. 181-205.

Almond, G. and S. Verba (1963), *The Civic Culture: Political Attitudes and Democracy in Five Nations*, Sage Publications, London.

Alsop, R., M. Bertelsen and J. Holland (2006), "Empowerment in practice: From analysis to implementation", *Directions in Development*, World Bank, Washington, DC.

Arnett, J.J. (2008), "The neglected 95%: Why American psychology needs to become less American", *American Psychologist,* Vol. 63, No. 7, pp. 602-614.

Arrow, K. (1972), "Gifts and exchanges", *Philosophy & Public Affairs,* Vol. 1, No. 4, pp. 343-362.

Benz, M. and S. Meier (2006), "Do people behave in experiments as in the field? Evidence from donations", *Federal Reserve Bank of Boston Working Papers,* No. 06-8.

Berg, J., J. Dickhaut and K. McCabe (1995), "Trust, reciprocity, and social history", *Games and Economic Behavior,* Vol. 10, pp. 122-142.

Boarini, R. et al. (2012), "What makes for a better life? The determinants of subjective well-being in OECD countries: Evidence from the Gallup World Poll", *OECD Statistics Working Papers,* No. 2012/03, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k9b9ltjm937-en*.

Boreham, R. et al. (2004), "Neighbourhood social cohesion and health: Investigating associations and possible mechanisms", *Social Capital for Health: Issues of definition, Measurement and Links to Health*, A. Morgan and C. Swann (eds.), Health Development Agency, London.

Bouckaert, G. (2012), "Trust and public administration", *Administration*, Vol. 60, No. 1, pp. 91-115.

Bouckaert, G. and S. Van de Walle (2003), "Comparing measures of citizen trust and user satisfaction as indicators of 'good governance': Difficulties in linking trust and satisfaction indicators", *International Review of Administrative Sciences*, Vol. 69, No. 3, pp. 329-343.

Bourdieu, P. (1984), *Distinction: A Social Critique of the Judgement of Taste*, Routledge & Keegan Paul (eds), London.

Capra, C.M., K. Lanier and S. Meer (2007), "Attitudinal and behavioral measures of trust: A new comparison", *Working Paper*, Department of Economics, Emory University, Atlanta.

Coleman, J. (1988), "Social capital in the creation of human capital", *American Journal of Sociology*, Vol. 94, Supplement: Organizations and Institutions: Sociological and Economic Approaches to the Analysis of Social Structure, pp. S95-S120.

Cook, T.E. and P. Gronke (2005), "The skeptical American: Revisiting the meanings of trust in government and confidence in institutions", *Journal of Politics, Vol.* 67, No. 3, pp. 784-803.

Cox, J. (2004), "How to identify trust and reciprocity", *Games and Economic Behaviour,* Vol. 46, pp. 260-281.

Dasgupta, N. and S. Asgari (2004), "Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping", *Journal of Experimental Social Psychology*, Vol. 40, No. 5, pp. 642-658.

Daude, C., H. Gutiérrez and A. Melguizo (2012), "What drives tax morale?", *OECD Development Centre Working Papers*, No. 315, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k8zk8m61kzq-en*.

Davidov, E. and L. Coromina (2013), "Evaluating measurement invariance for social and political trust in Western Europe over four measurement time points (2002-2008)", *ASK. Research and Methods,* Vol. 22, pp. 37-54.

Delhey, J., K. Newton and C. Welzel (2011), "How general is trust in 'most people'? Solving the radius of trust problem", *American Sociological Review,* Vol. 76, No. 5, pp. 786-807.

Eurobarometer (2016), *Eurobarometer Data Service,www.gesis.org/eurobarometer-data-service/search-data-access/data-access* (accessed 06 February 2017).

Eurofund (2016), *European Quality of Life Survey*, *www.eurofound.europa.eu/surveys/european-quality-of-life-surveys* (accessed 06 February 2017).

European Social Survey (2016), *European Social Survey, www.europeansocialsurvey.org/data/* (accessed 06 February 2017).

Eurostat (2016), *European Union Statistics on Income and Living Conditions (EU SILC)*, *http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions* (accessed 23 March 2017).

Falk, A. et al. (2016), "The preference survey module: A validated instrument for measuring risk, time, and social preferences", *Working Paper*, University of Bonn.

Falk, A. et al. (2015), "The nature and predictive power of preferences: Global evidence", *IZA Discussion Paper,* No. 9504.

Fehr, E. (2009), "On the economics and biology of trust", *Journal of the European Economic Association,* Vol. 7, No. 2-3, pp. 235-266.

Fehr, E. et al. (2003), "A nation-wide laboratory: Examining trust and trustworthiness by integrating behavioral experiments into representative surveys", *CESifo Working Paper Series,* No. 866; *IZA Discussion Paper,* No. 715.

Felte, E. (2001), "Finders keepers", *Reader's Digest,* 2 April.

Fleche, S., C. Smith and P. Sorsa (2012), "Exploring determinants of subjective well-being in OECD countries: Evidence from the WVS", *OECD Statistics Working Papers,* No. 2012/01, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k9ffc6p1rvb-en*.

Frey, B.S. and A. Stutzer (2006), "Political participation and procedural utility: An empirical study", *European Journal of Political Research*, Vol. 45, No. 3, pp. 391-418.

Frey, B.S. and A. Stutzer (2005), "Beyond outcomes: measuring procedural utility", *Oxford Economic Papers*, No. 57, No. 1, pp. 90-111.

Frey, B.S., M. Benz and A. Stutzer (2004), "Introducing procedural utility: Not only what, but also how matters", *Journal of Institutional and Theoretical Economics,* Vol. 160, No. 3, pp. 377-401.

Fukuyama, F. (1995), *Trust: The Social Virtues and the Creation of Prosperity*, Hamish Hamilton, London.

Gachter, S., B. Herrmann and C. Thoni (2004) "Trust, voluntary cooperation, and socioeconomic background: Survey and experimental evidence", *Journal of Economic Behavior & Organization*, Vol. 55, No. 4, pp. 505–531.

Gallup (2016), *Gallup World Poll, www.gallup.com/services/170945/world-poll.aspx* (accessed 06 February 2017).

Ginn, J. and S. Arber (2004), "Gender and the relationship between social capital and health", *Social Capital for Health: Issues of definition, Measurement and Links to Health*, A. Morgan and C. Swann (eds.) Health Development Agency, London.

Glaeser, E. et al. (2000), "Measuring trust", *The Quarterly Journal of Economics,* Vol. 115, No. 3, pp. 811-846.

Gonzalez, S. and C. Smith (2017), "The Accuracy of Measures of Institutional Trust in Household Surveys: Evidence from the OECD Trust Dataset", *OECD Statistics Working Papers,* OECD Publishing, Paris.

Greenwald, A.G., B.A. Nosek and M.R. Banaji (2002), "Understanding and using the implicit association test: I. An improved scoring algorithm", *Journal of Personality and Social Psychology*, Vol. 85, No. 2, pp. 197-216.

Guiso, L., P. Sapienza and L. Zingales (2008), "Trusting the stock market", *The Journal of Finance,* Vol. 63, No. 6, pp. 2557-2600.

Guiso, L., P. Sapienza and L. Zingales (2006), "Does culture affect economic outcomes?", *The Journal of Economic Perspectives,* Vol. 20, No. 2, pp. 23-48.

Hamilton, K. and G. Liu (2013), "Human capital, tangible wealth, and the intangible capital residual", *OECD Statistics Working Papers,* No. 2013/02, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k4840h633f7-en.*

Hardin, R. (2004), *Trust and Trustworthiness*, Russel Sage Foundation, New York.

Helliwell, J., H. Huang and S. Wang (2016), "The distribution of world happiness", *World Happiness Report 2016, Update (Vol. I)*, Helliwell, J., R. Layard and J. Sachs (eds), Sustainable Development Solutions Network, New York.

Helliwell, J. and S. Wang (2010), "Trust and well-being", *NBER Working Papers,* No. 15911, National Bureau of Economic Research.

Helliwell, J. and R. Putnam (2004), "The social context of well-being", *Philosophical Transactions of the Royal Society,* Vol. 359, pp. 14.

Henrich, J., S. Heine and A. Norenzayan (2010), "Most people are not WEIRD", *Nature,* Vol. 446, p. 29.

Intawan, C. and S.P. Nicholson (2017), "My trust in government is implicit: Automatic trust in government and system support", unpublished manuscript.

ISTAT (2013), *Benessere Equo e Sostenibile,* Rome.

Johnson, N. and A. Mislin (2012), "How much should we trust the WVS trust question?", *Economics Letters,* Vol. 116, pp. 210-212.

Johnson, N. and A. Mislin (2011), "Trust Games: A meta-analysis", *Journal of Economic Psychology,* Vol. 32, No. 5, pp. 865-889.

Karlan, Dean, S. (2005), "Using experimental economics to measure social capital and predict financial decisions", *American Economic Review*, Vol. 95, No. 5, pp. 1688-1699.

Kawachi, I. and L. Berkman (eds.), (2000), *Social Epidemiology*, Oxford University Press, New York.

Klijn, E.H., J. Edelenbos and B. Steijn (2010), "Trust in governance networks: Its impact and outcomes", *Administration and Society*, Vol. 42, No. 2, pp. 193-221.

Knack, S. (2001), "Trust, associational life, and economic performance", *MPRA Paper,* No. 27247, University of Munich.

Knack, S. and P. Keefer (1997), "Does social capital have an economic payoff? A cross-country investigation", *Quarterly Journal of Economics,* Vol. 112, No. 4, pp. 1251-88.

La Porta, R. et al. (1997), "Trust in large organisations", *American Economic Review, Papers and Proceedings,* Vol. 87, pp. 333-38.

Lazzarini, S.G. et al. (2004), "Measuring trust: An experiment in Brazil", *Insper Working Papers,* Vol. 049/2004.

Marien, S. and M. Hooghe (2011), "Does political trust matter? An empirical investigation into the relation between political trust and support for law compliance", *European Journal of Political Research,* Vol. 50, No. 2, pp. 267-291.

McKnight, D.H., V. Choudhury and C. Kacmar (2002), "The impact of initial consumer trust on intentions to transact with a web site: A trust-building model", *The Journal of Strategic Information Systems*, Vol. 11, No. 3, pp. 297-323.

Ministry for Environmental Protection, Israel (2014), *Well-being Indicators for Israel*, State of Israel, Jerusalem.

Morrone, A., N. Tontoranelli and G. Ranuzzi (2009), "How good is trust? Measuring trust and its role for the progress of societies", *OECD Statistics Working Papers,* No. 2009/03, OECD Publishing, Paris, *http://dx.doi.org/10.1787/220633873086*.

Murphy, K. (2004), "The role of trust in nurturing compliance: A study of accused tax avoiders", *Law and Human Behaviour*, Vol. 28, No. 2, pp. 187-209.

Murphy, K., T.R. Tyler and A. Curtis (2009), "Nurturing regulatory compliance: Is procedural justice effective when people question the legitimacy of the law?", *Regulation and Governance,* Vol. 3, No. 1, pp. 1-26.

Naef, M. and J. Schupp (2009), "Measuring trust: Experiments and surveys in contrast and combination", *IZA Discussion Paper,* No. 4087.

Nannestad, P. (2008), "What have we learned about generalized trust, if anything?", *Annual Review of Political Science,* Vol. 11, pp. 413-436.

New Zealand Ministry of Justice (2014), *New Zealand Crime and Safety Survey, www.justice.govt.nz/justice-sector-policy/research-data/nzcass/* (accessed 21 April 2017).

Nooteboom, B. (2007), "Social capital, institutions and trust", *Review of Social Economy,* Vol. 65, No. 1, pp. 29-53.

OECD (2017a), *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264268920-en*.

OECD (2017b), *Measuring Distance to the SDGs Targets: An Assessment of where OECD Countries Stand*, OECD Publishing, Paris, *www.oecd.org/std/OECD-Measuring-Distance-to-SDG-Targets.pdf*.

OECDStat (2016), *OECD Statistics, https://stats.oecd.org/* (accessed 22 April 2017).

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

OECD (2013a), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

OECD (2013b), *Government at a Glance*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/gov_glance-2013-en*.

OECD (2013c), *How's Life? 2013: Measuring Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264201392-en*.

OECD (2011), *How's Life*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264121164-en*.

OECD (2008), "Quality Framework and Guidelines for OECD Statistical Activities", OECD Publishing, Paris.

OECD (2001), *The Well-being of Nations*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264189515-en*.

ONS (2016), *Measuring National Well-being in the UK: Domains and Measures*, Office for National Statistics, UK, London.

Penczynski, S. and M. Santana (2016), *Measuring Trust in Institutions: A Lab-in-the-Field Study Using Time Preference Elicitation*, University of Mannheim.

Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, New Jersey.

Reeskens, T. and M. Hooghe (2008), "Cross-cultural measurement equivalence of generalized trust. Evidence from the European Social Survey, 2002 and 2004", *Social Indicators Research,* Vol. 85, pp. 515-532.

Rosenberg, M. (1957), "Misanthropy and attitudes toward international affairs", *Journal of Conflict Resolution*, pp. 340-345.

Roth, F. (2015), "Political economy of EMU. Rebuilding systemic trust in the euro area in times of crisis", *EC Discussion Papers, No.* 16, Brussels.

Rothstein, B. and E. Uslaner (2005), "All for all: Equality, corruption and social trust", *World Politics,* Vol. 58, No. 3, pp. 41-72.

Sampson, R. (2012), *Great American City: Chicago and the Enduring Neighbourhood Effect*, The University of Chicago Press, Chicago and London.

Schneider, I. (2016), "Can we trust measures of political trust? Assessing measurement equivalence in diverse regime types", *Social Indicators Research, http://dx.doi.org/10.1007/s11205-016-1400-8*.

Scrivens, K. and C. Smith (2013), "Four interpretations of social capital: An agenda for measurement", *OECD Statistics Working Papers,* No. 2013/06, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jzbcx010wmt-en*.

Sen, A.K. (2009), *The Idea of Justice,* Harvard University Press, Cambridge.

Sen, A.K. (1992), *Inequality Re-examined,* Harvard University Press, Cambridge.

Shanahan, E. (2007), "Excuse me, is this your phone?", *Reader's Digest,* July.

Smith, V.L. (2008), "Experimental economics", *The New Palgrave Dictionary of Economics*, 2nd Edition.

Stiglitz, J.E., A. Sen and J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, *www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf*.

Sturgis, P. and P. Smith (2010), "Assessing the validity and reliability of generalized trust questions: What kind of trust are we measuring?", *International Journal of Public Opinion Research,* Vol. 22, No. 1.

Transparency International (2016), *Corruptions Perception Index*, *www.transparency.org/research/cpi/overview* (accessed 06 February 2017).

UNECE (2014), *Conference of European Statisticians Recommendations on Measuring Sustainable Development*, United Nations, New York and Geneva.

United Nations (2014), *Fundamental Principles of Official Statistics*, No. A/RES/68/261, United Nations General Assembly, New York.

United Nations (2015), "Transforming our world: The 2030 Agenda for Sustainable Development", No. A/RES/70/1, United Nations General Assembly, New York.

UNODC (2016), *Crime and Criminal Justice,* United Nations Office on Drugs and Crime, *www.unodc.org/unodc/en/data-and-analysis/* (accessed 06 February 2017).

Uslaner, E. (2008), "Trust as a moral value", *The Handbook of Social Capital*, D. Castiglione, J. Van Deth and G. Wolleb (eds.), Oxford University Press, Oxford, pp. 101-121.

Uslaner, E. (2002), *The Moral Foundations of Trust*, Cambridge University Press.

Whitely, P. (1997), *Economic Growth and Social Capital*, Political Economy Research Centre, Sheffield.

World Bank (2006), *Where is the Wealth of Nations? Measuring Capital for the 21st Century*, The World Bank, Washington, DC.

Zak, P. and S. Knack (2001), "Trust and growth", *The Economic Journal,* Vol. 11, pp. 295-321.

Zelmer, J. (2003), "Linear public goods experiments: A meta-analysis", *Experimental Economics,* Vol. 6, pp. 299-310.

Zhao, H. and S. Kim (2011), "An exploratory examination of the social capital and FDI linkage and the moderating role of regulatory quality: A cross-country study", *Thunderbird International Business Review,* Vol. 53, No. 5, pp. 629-646.

© OECD 2017

Chapter 3

Methodological considerations

This chapter addresses potential measurement error in trust survey data, focusing on the way survey design can either exacerbate or mitigate it. For a range of issues (question wording, response formats, survey context, survey mode, cross-cultural response styles), evidence on the key methodological challenges for data quality and key practical messages for survey design are highlighted. The chapter concludes by pointing out areas where additional methodological research will be needed.

3.1. Introduction

This chapter addresses the potential for measurement error in survey data on trust and the way it interacts with – and can be influenced by – the survey design. Following a brief overview of the different types of response biases that respondents can exhibit and that can cause measurement error, the chapter discusses various methodological aspects of the design of questions and surveys that impact on these biases.

The chapter specifically discusses the methodological issues that are of particular importance for measuring trust. These include question wording (Section 3.3), response formats (Section 3.4), the survey context (Section 3.5), survey mode (Section 3.6) and (cross-cultural) response styles (Section 3.7). Each of these sections presents evidence on the key methodological challenges for data quality and highlights key messages for survey design. Wherever possible, the evidence is drawn from trust-specific studies. However, as the methodological literature on trust questions is quite meagre, broader evidence from self-reported measures is relied on heavily. In this context, the 2013 OECD Guidelines on the Measurement of Subjective Well-being provide a useful reference point for a much broader discussion of methodological considerations in measuring intangible concepts. The chapter concludes with directions for further research.

3.2. Measurement error

It is important to recognise that all measures, even objective ones, exhibit some degree of error. Hence, the goal is not to select a perfect (and probably non-existent) measure but rather one that is "good enough" to distinguish meaningful patterns, such as changes in trust over time and differences between population subgroups, from noise in the data. As trust items are relatively sensitive to varying survey conditions and to how questions are framed, advice on their measurement needs to be more specific than is the case for some more "objective" indicators, such as educational attainment. Arguably, however, this sensitivity also exists in many other self-reported survey measures that are already being collected, such as subjective well-being measures or income.

Measurement error is the extent to which survey measures reflect concepts other than those intended by the surveyor (OECD, 2013). This error can be either systematic, leading to a bias in the data that is consistent in some way, and that might result in, for example, values that are consistently higher or lower than might be expected, or random, i.e. varying in an unpredictable manner (Maggino, 2009). The risk of error is essentially the product of a complex interaction between methodological factors (such as the cognitive demands made by certain questions, or the contextual features of a survey that might influence responses), respondent factors (such as motivation, fatigue and memory), and the construct of interest itself (such as how interesting or relevant respondents find the survey).

In order to answer any survey question, respondents are assumed to go through several cognitive steps, which may be performed either sequentially or in parallel. These steps include understanding the question, recalling information from memory, forming a judgement,

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

formatting the judgement to fit the response alternatives, and editing the final answer before delivering it to the surveyor (Sudman, Bradburn and Schwarz, 1996). It is important to understand this question-answer process not as a simple robotic task, but as part of a social interaction process between respondent, interviewer, question design and the survey context.

Aspects of survey design and context can either cause or exacerbate measurement error. Respondent failures in memory, motivation, communication or knowledge, which all can lead to respondent error in self-reported measures, are often associated with an increased risk of response biases and the use of response heuristics (Bradburn, Sudman and Wansink, 2004). Response biases refer to particular patterns or distortions in how individuals or groups of individuals respond to questions, while response heuristics refer to (often sub-conscious) mental shortcuts that respondents rely on to choose their answers (OECD, 2013). Drawing on the classifications of Podsakoff et al. (2003), Table 3.1 provides an overview of the response biases and heuristics commonly associated with all self-reported measures (OECD, 2013). Some, but not all, can apply to trust measures. The following sections consider the various methodological features of survey design that can lead to these most relevant response biases.

Table 3.1. **Overview of response biases and heuristics**

Response bias or heuristic	Exhibited response pattern
Acquiescence or yea-saying	A tendency to agree with, or respond positively to, survey items regardless of their content.
Nay-saying	A tendency to disagree with, or respond negatively to, survey items regardless of their content.
Extreme responding	A tendency to use response categories towards the ends of a response scale/the most extreme response category.
Moderate responding	A tendency to use responses towards the middle of the response scale/the most moderate response category.
No-opinion responding	A tendency to select the response category that is most neutral in its meaning (e.g. neither agree nor disagree).
Random responding	A tendency to respond randomly, rather than meaningfully.
Digit preferences	On numerical response formats, a tendency to prefer using some numbers more than others.
Primacy effects	A tendency to select one of the first response categories presented on a list.
Recency effects	A tendency to select one of the last response categories presented on a list.
Socially desirable responding	Conscious or subconscious tendency to select response options more likely to conform with social norms or present the respondent in a good light.
Demand characteristics	A reaction to subtle cues that might reflect the surveyor's beliefs about how they should respond and/or their own beliefs about the purpose of the survey (e.g. "leading questions", where the tone or phrasing of the question suggests to respondents that particular answers should be favoured).
Consistency motif or bias	A tendency for respondents to try and ensure consistency between responses (e.g. consistency between a question about attitudes towards smoking and a question about cigarette purchasing habits).
Priming effects	Where the survey context (e.g. question order; survey source) influences how questions are understood or makes certain information more easily accessible to respondents.

Source: OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

StatLink *http://dx.doi.org/10.1787/888933584146*

3.3. Question wording

Evaluations of the possible effects of question construction are central to guaranteeing the comparability, internal consistency and test-retest reliability of seemingly similar yet slightly different survey items. Questions that are easily understood and not ambiguous and that do not pose an unnecessary burden on respondents will increase the validity of responses and at the same time depress error variability (OECD, 2013).

For trust measures, question wording encompasses aspects of question comprehension, translatability across languages and societal subgroups, and changes in wording. This section reviews the challenges that can arise from these aspects, with a particular focus on the effect of question wording (as most available studies focus on this aspect).

The issues

First, if we want to compare self-reported items of trust across survey participants and countries and reduce variability due to measurement error, it is essential that respondents comprehend and interpret questions in a similar and unambiguous way. This also includes translatability across languages between countries and across different socio-economic and demographic subgroups within a society. For example, survey comparability is compromised if a certain wording of trust questions evokes different connotations for old vs. young people or does not conceptually exist in certain cultures.

Second, an important question in survey design, particularly for attitudinal questions like trust measures, relates to whether micro-changes in question wording significantly influence results. On the one hand, it would be worrying if changes in question wording do not produce different results, as this would imply that respondents do not understand question nuances well and do not cognitively process the concepts that the different questions are trying to distinguish between. On the other hand, it can be problematic if a slightly different wording of items that have been designed to measure the same underlying construct (and are often used interchangeably in comparative research) actually leads respondents to interpret the questions in dissimilar ways.

The evidence

A common way to evaluate question comprehension (apart from cognitive testing, the results of which are rarely published) is to look at an item's response latencies, or the time taken to process a question and deliver an answer. However, it is not absolutely clear whether short response latencies for an item indicate whether it was well understood or whether the respondent answered randomly. Unfortunately, in either case no study has so far considered response speed with regard to trust questions. The best evidence that respondents have understood question meaning and provided meaningful answers is demonstrated by strong correlations between the measures themselves and real-world outcomes as well as other non-survey indicators (OECD, 2013). As highlighted in the previous chapter, considerable support does exist for the overall validity of trust measures, particularly with regards to interpersonal trust.

We now turn to the issue of question wording. Two central aspects will be considered in the following – how similar question wording has to be to capture the same underlying concept, as well as how specific question wording has to be to conclude that it indeed taps into trust.

When it comes to evaluative measures of interpersonal trust, most items that can be found across surveys and time do not very greatly depart from the original Rosenberg question: "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?" Nevertheless, a few surveys, sometimes routinely and sometimes as part of intentional methodological research, have featured different versions of the original trust question. Although not all studies are split-sample experiments that would allow for the definite attribution of different response distributions to changes in question phrasing, the available evidence suggests that responses to interpersonal trust measures are quite sensitive to wording changes.

For example, Smith (1997) capitalises on the fact that the American General Social Survey (GSS), which has featured the Rosenberg scale as a core battery item since 1972, has additionally included variants of the standard question over the years. An examination of

these shows quite different response patterns between versions. In a 1983 GSS experiment, 57% of respondents answered *yes* to the question "do you think that most people can be trusted?", while only 36.5% indicated that *most people can be trusted* when the item was phrased "Some people say that most people can be trusted. Others say you can't be too careful in your dealing(s) with people. How do you feel about it?" Both items offered a dichotomous answering scale. A drop like this of 20.5% of trusting respondents between question versions is quite extraordinary. As Smith himself notes, this difference might be driven not only by question wording effects but also by response scale and order effects, as the items feature slightly different response options (a simple *yes/no* vs. a more explicit *most people can be trusted/you can't be too careful*). Further, both items were preceded by questions on quite different other topics. Response scale type and question order within the survey context are potentially large sources of error variability, matters which are discussed in Sections 3.4 and 3.5. But even just considering the known impacts of question construction documented in the literature on attitudinal questions more generally, a different response distribution and the direction of difference between the two questions intuitively makes sense: on the one hand, *do you think that most people can be trusted?* is an unbalanced question that only specifies one direction of trust, instead of also spelling out the alternative *or do you think that most people cannot be trusted?* For other attitudinal questions beyond trust, which also commonly refer to intangible and ambiguous concepts, such balance effects have often been associated with encouraging acquiescence (Schuman and Presser, 1996; Peabody, 1961). On the other hand, the second question version, just like the standard Rosenberg question, includes a reference to *being careful*, which one could argue is not quite the same as distrust and thus introduces a different underlying concept to the respondent.

It has been recognised in other places that the concept of caution in the trust question might be problematic, since being careful could carry quite different connotations for different population subgroups. Hooghe and Ressken (2008) write that "carefulness might imply something else for someone who is weak and vulnerable, compared to an athletic, bright and well-off person."

Empirical support for this hypothesis has been offered by at least two studies: Soroka, Helliwell and Johnston (2007) examine four different versions of interpersonal trust questions included in the 2000/2001 Equality, Security and Community (ESC) survey carried out in Canada,[1] two of which feature a "caution rider". First, the authors also find evidence of the above-mentioned balance and acquiescence effects, depending on whether the question is offering a full balance of trust/non-trust alternatives. Second and more importantly, the authors assert that "saying that you cannot be too careful in dealing with people is not the same as saying that people in general cannot be trusted. The first represents a cautious disposition, while the latter simply represents the reverse of the question 'would you say that most people can be trusted?'" (ibid., p. 114) They further find that while women are less trusting than men when the standard (balanced) trust question is used, they are more trusting than men when a question without the *cannot be too careful* rider is used. It therefore may be cautiousness, rather than trustworthiness, that drives gender differences in the Rosenberg trust question.

The second piece of supporting evidence comes from two methodological experiments using split samples that the Office for National Statistics (ONS) in the United Kingdom carried out in co-operation with the OECD in October 2015 and May 2016 during its standard Opinions Survey. As each split sample in this experiment featured a relatively small sample size of 500, one cannot interpret the following results as absolutely definite – they are not

statistically significant unless explicitly stated so. Nevertheless, they confirm the intuition of a caution rider effect on certain groups that came out elsewhere. In the first experiment (see Table 3.2), respondents were offered either the standard Rosenberg question with a caution rider: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (with a dichotomous answering choice) or the more neutral 11-point scale European Social Survey version: "On a scale where 0 is not at all and 10 is completely, in general how much do you think people can be trusted?"

Table 3.2. **Comparison of interpersonal trust questions by gender and age groups**

Experimental round	Question wording	Response scale	Population				
			All	Men	Women	16-44 years	Above 45 years
October 2015	"Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"	Dichotomous	32.1	33.6	30.7	31.8	32.3
	"On a scale where 0 is not at all and 10 is completely, in general how much do you think people can be trusted?"	11-point scale	38.6	36.8	40.4	33.7*	43.1*
May 2016	"Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?"	Dichotomous	35.6	37.6	33.7	31.4*	40*
	"Generally speaking, would you say that most people can be trusted, or that you cannot be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you cannot be too careful and 10 means that most people can be trusted."	11-point scale	36.5	38.4	34.8	38.3	35.3

Note: For the questions using a dichotomous response scale, numbers indicate the proportion of the population indicating trust. For the questions using a 11-point response scale, numbers indicate the proportion of the population indicating a response between 7 and 10.
* denotes significance at the 10% level.
Source: ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, Oct 2015 to May 2016", Office for National Statistics, UK, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

StatLink http://dx.doi.org/10.1787/888933584165

When asked about trust with the caution rider version, fewer women reported that most people can be trusted than was the case for men (30.7% vs. 33.6%). By contrast, when using the question wording without the caution rider, more women (40.4%) reported a score of 7 to 10 than was the case for men (36.8%). A comparable pattern can be observed for older people (over 45 years) vs. younger people (aged 16-44): although a similar proportion of those aged 16-44 reported high levels of trust under both question versions, the same was not true for the older age group who, like women, reported lower levels of trust when the caution phrasing was present in the question.

In a follow-up experiment (also portrayed in Table 3.2), the phrasing of the 11-point scale item was changed slightly to also include a caution rider: "Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? Please tell me on a score of 0 to 10, where 0 means you can't be too careful and 10 means that most people can be trusted". This effectively reversed the results of the first experiment with regards to the 11-point scale. Whereas before women and older people were more likely to report higher levels of trust than the total population, with the caution rider now included, women (34.8%) and older people (35.9%) reported lower levels of trust than the total population (36.5%). Older people and women can arguably be considered as what Hooghe and Reeskens termed relatively more "vulnerable" to the actions of other

people.[2] It therefore seems plausible to conclude that interpersonal trust questions that use a *cannot be too careful* phrasing, compared to more neutral question wording that focuses solely on trust, induce a priming effect on relatively vulnerable groups. Resulting responses might reflect differences in cautiousness rather than trust.

When it comes to institutional trust, almost no studies have addressed issues of different question wording. A notable exception comes from the same ONS experiment in 2015/16 that has been described above. Here, two issues have been tested: one regarding whether specifying the context in which trust occurs makes a difference, and one regarding whether the word *trust* can be used interchangeably with the word *confidence*.

In the first experiment (see Table 3.3), respondents were presented with an *A trusts B* type question "I am going to name a number of organisations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?" vs. an **A** trusts **B** to do **X** type question "I am going to name a number of organisations. For each one, could you tell me how much confidence you have in them *to act in the national interest*: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?". The questions were applied to

Table 3.3. **Comparison of confidence in institutions versus confidence in institutions to act in the national interest**

Institution		Question wording	
		Confidence (%)	Confidence to act in the national interest (%)
Armed Forces			
	A great deal	54.6	59.7
	Quite a lot	37.3	31.7
Police			
	A great deal	22.1*	28.5*
	Quite a lot	50.9	47.5
Justice system			
	A great deal	10.7	14.3
	Quite a lot	44.8	41.2
Parliament			
	A great deal	3.8	6.2
	Quite a lot	25.4	24.9
Civil service			
	A great deal	9.2	12.4
	Quite a lot	51.6	45.8
National health service			
	A great deal	36.7	43.7
	Quite a lot	46.8	40.5
Banks			
	A great deal	8.5	5.5
	Quite a lot	39.8	26.7
	Combined	48.3**	32.2**
Media			
	A great deal	2.3	4.6
	Quite a lot	20	18.9

Note: Responses recorded on a 4-point scale: "a great deal", "quite a lot", "not very much" or "not at all".
* denotes significance at the 10% level.
** denotes significance at the 1% level.
Source: ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, Oct 2015 to May 2016", Office for National Statistics, UK, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

StatLink http://dx.doi.org/10.1787/888933584184

a range of different institutions, namely the armed forces, the police, the justice system, Parliament, the civil service, the National Health Service, banks and the media. The results, displaying different response distributions between the two question versions depending on the type of institution, suggest that adding a specific trust context can lead to a slightly different interpretation of the trust measure: adding *to act in the national interest* results in a higher proportion of respondents indicating *a great deal of confidence*, the highest possible category, for all tested institutions except the banks, compared to the unspecified question version. In stark contrast, only 5.5% of respondents indicated *a great deal of confidence* in the banks *to act in the national interest* version (compared to 8.5% in the unspecified question). This trend was even more obvious when considering the proportion of respondents indicating *quite a lot of confidence* in the banks: while 39.8% selected this choice for the unspecified question version, only 26.7% did so with the *to act in the national interest* item. It is also worth noting that the banks are the only institution for which both of these categories declined.

It could sensibly be argued that citizens may very well be relatively confident in banks to manage their money, but very much less likely to believe that financial institutions, in view of the financial crisis that affected the UK sample, act in the long-term interest of the country. This demonstrates that, at least for banks in a context of financial crisis, there is indeed a conceptual gap between *confidence* vs. *confidence to act in the national interest* for respondents and that this can make a difference to response distributions. It is not clear yet whether *to act in the national interest* is the best specification (or indeed better than no specification at all), and the impact of possible alternatives (e.g. *to act in my personal interest* or *to do what is right*) on the different types of institutions should be empirically examined in the future. It will also be important to test such specifications in other cultural contexts – the current results might apply to the UK only.[3]

In a second experiment, the ONS explored the distinction between the concepts of *confidence* vs. *trust* in institutions, with half the respondents being asked: "I am going to name a number of organisations. "For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all?" and the other half being presented with the question: "I am going to name a number of organisations. For each one, could you tell me how much you trust that institution: is it a great deal of trust, quite a lot of trust, not very much trust or none at all?" Although the theoretical literature on trust in institutions has suggested that confidence and trust tap into slightly different concepts (see Roberts and Hough, 2005 who claim that trust is something one does and is more concrete, whereas confidence is something one has and is more abstract), the experiment found no clear-cut evidence that this distinction is mirrored in how respondents actually respond to questions. As Table 3.4 shows, no consistent logical pattern between the two question versions appears – while it might be somewhat comprehensible that trust is lower than confidence in the media, it does not seem intuitively apparent why the justice system is the only institution for which respondents report higher trust than confidence. Moreover, if trust and confidence really capture two distinct concepts, it seems quite implausible that the banks, which have exhibited quite polarising results in the *confidence* vs. *confidence to act in the national interest* testing described earlier, have a similar proportion of respondents indicating *trust* and *confidence*.

It thus seems that, when distinctions between two concepts are too narrow, respondents cannot draw a meaningful separation between them. The issue of confidence vs. trust touches upon cultural comparability as well. For internationally valid measures, it is important to keep descriptors broad enough to be easily translatable (OECD, 2013). In

Table 3.4. **Comparison of confidence in institutions versus trust in institutions**

Institution		Question wording	
		Confidence (%)	Trust (%)
Armed Forces			
	A great deal	58.5	53.7
	Quite a lot	36.6	39.8
	Total	95.2	93.5
Police			
	A great deal	34.3	30.2
	Quite a lot	46.4	50.9
	Total	80.7	81.1
Justice system			
	A great deal	12.8	15
	Quite a lot	45	51.1
	Total	57.8	66.2
Parliament			
	A great deal	4.1	1.8
	Quite a lot	24.8	25.5
	Total	28.9	27.2
Civil service			
	A great deal	11.2	8.8
	Quite a lot	54	55
	Total	65.2	63.8
National health service			
	A great deal	41.3	38.5
	Quite a lot	45.3	47.8
	Total	86.7	86.3
Banks			
	A great deal	9.1	9.3
	Quite a lot	44.9	45.5
	Total	54	54.8
Media			
	A great deal	5	1.9
	Quite a lot	23.8	16.3
	Total	28.8	18.2

Note: Responses recorded on a 4-point scale: "a great deal", "quite a lot", "not very much" or "not at all".
Source: ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, Oct 2015 to May 2016", Office for National Statistics, UK, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

StatLink http://dx.doi.org/10.1787/888933584203

contrast to Anglophone countries, most other cultures do not even distinguish between the two terms of trust and confidence. For instance, the French, Spanish and German languages have only one word for trust (*confiance*, *confianza* and *Vertrauen*, respectively). A similar point regarding the distinctions between different types of trust (in different types of people and different institutions) was made in Chapter 2: while a few distinct sub-dimensions within each of the two main categories of interpersonal and institutional trust are identified by respondents, many of the even finer distinctions made between different types of trust are not very empirically informative.

Key messages

- There is convincing evidence that respondents understand questions about trust measures well, mainly based on their strong validity with regard to real world outcomes. This is especially true for interpersonal trust.

- The exact question wording matters for trust measures. Questions should choose wording that does not refer to concepts other than trust and that is specific and precise to the situation of interest. Regardless of the eventual approach to the question wording adopted, standardisation is key to ensure comparison over time and between groups/countries.
- For interpersonal trust, the standard Rosenberg question's introduction of the caution concept impacts on the distribution of responses, with more vulnerable groups (e.g. women and the elderly) reporting lower trust with this phrasing. Hence, a more neutral question wording is preferable for interpersonal trust.
- For institutional trust, specifying the context of institutional behaviour can make a difference in some cases. For example, in the case of financial institutions in contexts which were affected by the financial crisis, adding the specification *to act in the national interest* significantly impacts on respondents' evaluations. When planning other trust questions, it is worth further investigating which other specifications (e.g. *to act in my personal interest* or *to do what is right*) matter for which institutions, and whether a specification should be used at all.
- If concepts are too narrowly related, respondents might have difficulty differentiating between them, as is shown by the analysis of the phrasing of *trust* vs. *confidence* in institutions in the ONS sample. Question wording should be precise enough to be understood by respondents, without getting into extremely subtle nuances (which might also pose an issue for translatability across countries).

3.4. Response formats

The response format chosen for attitudinal measures such as trust questions can have a non-trivial effect on the validity, reliability and comparability of responses. The format needs not only to properly represent the construct of interest and represent the full range of possible responses (including no-response options) but also to be understandable to respondents in the sense that they can provide meaningful and consistent replies. Aspects of the response format considered in the following section include the scale length, scale labelling and response order. Where possible, evidence specific to trust questions will be drawn upon, otherwise lessons from the broader literature on attitudinal questions are emphasised.

The issues

Scale design deals with the maximisation of discriminatory power or sensitivity based on choosing an optimal range of response options for the concept at hand. A good scale will capture as much meaningful variation between responses as exists (which is especially important for single-item measures). Offering too few response options might lead to an inability to detect variation and frustrate respondents, who might feel their attitudes cannot be accurately expressed. However, respondents might be equally burdened by too many response categories, especially if the categories are too close for them to distinguish between cognitively.

Scale labelling, or the way in which response options are described, can influence responses by setting a reference frame about the expected range within which responses can fall. The choice of scale labels, including the anchors at the scale ends, therefore needs to reflect the full range of possible response categories, without compromising reliability by offering meaningless response options, which can lead respondents to respond randomly. Lastly, especially in combination with the selected survey mode, the order of the response options offered can lead to certain options being selected by default by satisficing respondents.

The evidence

When it comes to determining the type of scale and the optimal number of response options being offered to respondents, a couple of general rules have been established in the broader literature on self-reported measurements. For example, if all response categories have verbal labels, five categories are thought to be the maximum number that respondents can process without visual cues (Bradburn et al., 2004). If the response categories are numerical and anchored by verbal descriptions at the two scale ends, respondents tend to respond more successfully to longer scales, as only the end-points need to be memorised. Here, psychometric evidence suggests that the majority of respondents can reliably distinguish between more than six or seven categories (ibid.). In general, longer numerical scales (with verbally-labelled anchors at the ends) have been found to enhance test-retest reliability and internal consistency among attitude measures (Weng, 2004; Alwin and Krosnick, 1991; Scherpenzeel and Saris, 1997; Cummins and Gullone, 2000), and validity is likely to improve with an increasing number of numerical scale points (Preston and Colman, 2000). By the same token, Preston and Colman (2000) have also found the internal consistency and discriminatory power of scales with few items to be low in general. Another body of research has posed the question whether to use an odd or even number of response categories, the former including a natural scale mid-point. Chang (1994) argues that – while a mid-point can result in unmotivated or undecided respondents clustering their replies around the middle category – the absence of a mid-point forces respondents to declare an attitudinal preference, even where one might not exist, hence resulting in random responses. In the case of bipolar attitudes (*yes/no* or *agree/disagree*), several studies support an odd number of response options to provide a mid-point that gives respondents the ability to express a neutral position (Alwin and Krosnick, 1991; Bradburn et al., 2004).

The available evidence on aspects of scale labelling suggests that scale anchors can encourage specific response biases. For instance, the use of *agree/disagree* or *yes/no* has been linked to acquiescence bias or "yea-saying" regardless of the actual question content, or even to encouraging socially desirable responding (Krosnick, 1999). Further, it appears to be sensible to adopt absolute scale anchors (e.g. *completely/not at all*) to clearly mark scale ends and offer the full spectrum of possible attitudes to respondents (OECD, 2013). In terms of labelling the points along a scale, there is some debate about whether this should be done numerically or verbally. On the one hand, a few studies have argued that adding verbal labels to all numbered response categories can produce more reliable and stable responding by clarifying the meaning of the different categories to respondents (see Alwin and Krosnick, 1991; Pudney, 2010). On the other hand, contradicting evidence suggests that numerical scales are more accurate than verbally labelled ones and can help to convey scale regularity and equally spaced intervals to respondents (Newstead and Arnold, 1989; Maggino, 2009). Furthermore, numerical labelling allows for the use of a greater number of response options, which is of particular importance for single-item measures. Lastly, in the interest of international comparability, numerical scales are much less likely to pose translation challenges across different languages and contexts (OECD, 2013). Overall, as scale labelling can have a non-negligible impact on the distribution of responses, it will be essential to be consistent in the approach to labelling once one is adopted.

Not only can the number of response options and their labelling characteristics influence answers, but so can the order in which they are presented. Response order effects can be categorised as primacy effects (where satisficing respondents are more likely to select the earlier response options in lists) and recency effects (where respondents tend to choose

the later response options). While primacy effects tend to occur more when response options are presented visually (which can be a useful way to guide respondents), recency effects are more common when choices are read aloud by interviewers (Krosnick, 1999). Both types of order effect become stronger due to item difficulty and respondent fatigue, and respondents with low cognitive abilities are more prone to exhibit this type of response bias (ibid.). In general, these effects are more likely to pose a problem for scales that are fully verbally-labelled rather than numerical. However, even numerical response order should be presented consistently (i.e. 0-10 instead of 10-0) in order to minimise mental switching between positive and negative normative outcomes (OECD, 2013).

How does some of this generally applicable advice translate to trust measures? Until recently, the most commonly used measure of interpersonal trust – the Rosenberg question – has been dichotomous. This measure has been used in a large number of surveys, most notably the Civic Culture surveys, the American National Election Studies from 1964 to 1976 and from 1992 to the present, the General Social Survey from 1972 to the present, and the World Values Surveys (WVS) since 1980. In the late 1990s, the Swiss Household Panel and the Citizenship, Involvement and Democracy (CID) cross-national surveys in Europe shifted to a 0-10 numerical point scale. Shortly thereafter the European Social Survey (ESS) adopted the 11-point scale as well.[4]

If one considers the scale length that respondents themselves seem to prefer, Lundasen (2010) conducted qualitative pretesting of World Values Survey interpersonal trust questions on a small Swedish sample where he encouraged the respondents to think aloud while considering the items. The respondents expressed difficulty in grading items on scales with four different points or more and tended to prefer questions with dichotomous answers. However, since the sample in this study was extremely small, it is doubtful how much weight should be given to its findings, especially in light of the larger quantitative research available on the performance of the two types of scales.

On the one hand, Uslaner, while acknowledging that dichotomous trust measures are not without problems, states that the dichotomous question is well understood when people are asked what it means, and it is stable both over time when asked in panels and from parents to children (Uslaner, 2002, pp. 68-74). He also holds that while the arguments for the greater precision of the 11-point numerical scale used in the ESS and CID for both institutional and interpersonal trust seem compelling, respondents are confused by and unable to process the larger number of response options (Uslaner, 2009). Using the ESS and CID surveys for the US, Romania, Moldova and Spain (the last three of which include both dichotomous and 11-point-scale trust measures in institutions and other people), he finds what he calls “evidence of a flight to the middle” or “systematic clumping”. More than 40 per cent of all answers to the 11-point scale questions across the multiple surveys he considers concentrate around the middle of the distribution, namely around the 4 to 6 values. However, it is very debatable whether this response pattern can actually be interpreted as problematic. There is no empirical evidence that middle values do not actually reflect how respondents feel. On the contrary, items that do not include mid-points, such as the dichotomous measure, might actually force respondents to express an attitude that does not reflect their true preference.

On the other hand, at least three studies have explicitly spoken in favour of the 11-point scale over the dichotomous scale for trust measures. Hooghe and Reeskens (2008) analysed data from the 2006 Belgian Youth Survey, which included both measures of

interpersonal trust (with about 40 questions between both items to prevent carry-over effects). Using regression analysis with respondents' demographic data, they found that the 11-point measure of trust, unlike the dichotomous measure, predicts involvement in voluntary associations, which has been linked to trust theoretically in the social capital literature (see Stolle, [1998. However, these findings need to be interpreted in the light of the sample consisting of 16-year-olds and the authors using a combined multi-item measure of trust and fairness (as envisaged by the original Rosenberg scale). Zmerli and Newton (2008) argue that there are theoretical reasons to assume that interpersonal and institutional trust go together, yet many studies find no significant correlations between the two constructs. The authors argue that this is because the majority of this research is based on surveys with short trust scales (4-point scales or dichotomies), such as the World Values Survey and the Eurobarometer. In fact, three of the four cross-sectional studies that find significant correlations between interpersonal and institutional trust use 11-point rating scales (Jagodzinski and Manabe, 2004; Denters, Gabriel and Torcal, 2007; Zmerli, Newton and Montero, 2007). Drawing on 24-country data from the ESS and US CID surveys that include the three Rosenberg interpersonal trust questions as well as six institutional trust questions (all on an 11-point scale), Zmerli and Newton themselves find a strong and robust correlation between interpersonal trust and confidence in institutions after controlling for various socio-economic factors that have commonly been found to be associated with the two types of trust. The results stand even when the three-variable Rosenberg scale is replaced with the single 11-point measure for generalised trust. Saris and Gallhofer (2006) come to similar conclusions when examining the British pilot study of the second wave of the ESS survey, which asked the same questions about interpersonal trust and institutional trust twice, once with a 4-point rating scale and once with an 11-point scale. In both cases, the three questions of the Rosenberg scale were used to measure interpersonal trust, and three questions about parliament, the legal system and the police were asked to tap confidence in public institutions. Only one of the correlations between interpersonal trust and institutional confidence was statistically significant in the case of the 4-point scale, but all nine were significant for the 11-point scale. Neither study, by Saris and Gallhofer or by Zmerli and Newton, addresses whether the correlations found could be due to shared method variance, and further research in this area will be needed to draw definite conclusions.

Key messages

- Different response options lead to different and not necessarily interchangeable measures. Therefore, a standardised approach to response format to ensure the consistency of measurement, especially in an international context, is highly advised.
- The available evidence in terms of general studies and specific information from trust measures suggests that a numerical 11-point scale with verbal scale anchors is preferable over the alternatives, as it allows for a greater degree of variance in responses and increases overall data quality as well as translatability across languages.
- Numerical response order should be presented consistently (i.e. 0-10 instead of 10-0) in order to minimise mental switching between positive and negative normative outcomes.
- When choosing scale anchors, the labels should represent absolute responses (e.g. completely/not at all) to minimise acquiescence bias and socially desirable responding and to allow for the full spectrum of possible responses.

3.5. Survey context

Results can be influenced not only by the design of the individual item itself, but also by the wider survey context in which the item is situated. Since individual survey questions are not asked in isolation, but as part of a continuous flow of items, the position within a sequence of items may conceivably influence responses. Apart from the question order within both a module and the larger survey, the broader temporal context in which the survey takes place is another defining feature of the survey context. While the survey mode, or the way in which data is collected, can also be categorised under survey context, this aspect will be addressed separately in the next section.

The issues

Since due to the very nature of language, words and sentences take part of the meaning from the environment in which they occur (Searle, 1979), the survey context can potentially influence how respondents understand and contextualise individual questions, as well as which information is brought to their minds when forming the answers to items. This can include the way questions or modules are introduced, as well as the nature of the questions asked immediately before trust measures. Attitudinal measures have been linked to respondents being likely to construct answers on the spot (as opposed to systematically retrieving specific memories), making them especially prone to context effects (Sudman, Bradburn and Schwarz, 1996; Schwarz and Strack, 2003).

For example, if questions about (potentially negative) experiences with other people are asked just before interpersonal trust questions, or questions about a specific type of government service are posed before general institutional trust items, a respondent might misinterpret what is intended by the actual trust question. Either consciously or sub-consciously, he or she might answer a question mainly in relation to the subject of the preceding items rather than based on the intended focus for the question, an effect often referred to as priming. Priming can influence both the mean level of a measure and the distribution of data if it impacts subgroups within a population differently, hence impairing comparability across surveys and between groups within the same survey. Further, for trend studies that are interested mainly in marginal changes, order effects are important even if they were to shift results only slightly. Depending on the nature of the priming, order effects can also suppress or inflate correlations between the variables, therefore putting substantive conclusions about their relationships in doubt.

The placement of trust questions within the larger survey context thus can strongly influence which information respondents take into account when constructing their answers, and order effects should be kept in mind when designing questionnaires. Apart from priming through the survey content itself, what is in the top of their minds when respondents formulate answers on their trust level could also be affected by the larger context in which a survey takes place, such as amidst major political scandals or events that could infer perceptions of interpersonal trust, such as a terrorist attack or the exposure of fraud schemes.

The evidence

The literature on the effect of question order on responses distinguishes between two possible directions of influence that can occur, namely assimilation and contrast effects. Assimilation effects are thought to occur when responses are consistent with the information being made salient through a priming question, while contrast effects take place when a response contrasts with the previous information (OECD, 2013).

Schuman and Presser (1996) point out that while order effects for attitudinal questions are not a rare occurrence, it is not the case that *any* disruption of sequence changes responses, and due to the possibility of either assimilation or contrast effects occurring, order effects can actually be hard to predict. Merely placing two items with similar content next to each other does not necessarily create an order effect. Only if respondents have a need to make their answer to the second question consistent with their answers to the first will such an order effect be created. Therefore, more targeted methodological research will be needed to discover what type of question or what type of context triggers which effect. However, a couple of general suggestions about how to deal with question order effects can be derived from experiments across 34 different US surveys (e.g. the Detroit Area Study and the NORC General Social Survey) in the 1970s. Firstly, context effects occur most often when two or more questions deal with the same or closely related issues (ibid.). It thus seems logical that they could be prevented if such items are separated within the questionnaire and are buffered by either intervening text or questions. Deaton (2011) found that the influence of political questions on life evaluations was reduced when a "buffer" question between institutional questions and life evaluations was added. More systematic research on the actual impact of transitional text or questions will be needed, as their effect will depend on whether the buffer items are related or unrelated to the actual questions of interest (Schwarz and Schuman, 1997). Furthermore, since the smooth organisation of a survey usually groups similar items together for reasons of coherence and to reduce the response burden, a compromise between avoiding order effects and maximising the ease of interviewing will have to be struck. The second general rule that Schuman and Presser establish states that general summary-type questions that ask respondents to evaluate complex topics in an overall way seem to be more sensitive to position than are more specific questions. This speaks for moving from a broad to a narrow level of specificity, e.g. by placing items about general interpersonal trust before questions about trust in specific groups, or starting with trust in government before asking about specific institutions.

Only very few studies have dealt with order effects in the specific context of trust measurements. In the case of institutional trust, Smith (1981) used a split-sample experiment to test the impact of a set of political alienation questions on subsequent institutional trust items and found that only the very first item (on confidence in major companies) showed a significant decline in trusting responses. This might be an indication of what is called a salience effect, where the first item in a list is more likely to be affected by order effects. Such salience effects will have to be taken into account whenever a list of items is used, for example in the case of institutional trust modules that ask about various organisations. This finding also highlights the need for consistency of order within question modules across surveys and over time.

In the case of interpersonal trust, Smith (1997), who examines the Rosenberg scale (which includes the general interpersonal trust question alongside items on the fairness of other people and the likelihood of their taking advantage of the respondent) in the US General Social Survey, states that "these items are prone to context effects because they call for global assessments of people in general based presumably on one's entire life experience. Making judgements based on such massive, cognitive retrievals are difficult and open to variability. Sampling of one's own memories on such broad topics tend to be biased rather than complete or random. Questionnaire context is one factor that biases the cognitive processing and in turn influences the summary judgments" (p. 174). Specifically, Smith found that trusting responses decline by 7.7% when the question was preceded by

items on negative experiences (crime and victimisation) rather than by questions on political ideology, the equalisation of wealth, divorce laws or the legalisation of marijuana.

Very interestingly, the direction of influence between attitudinal and experience measures can also run the opposite way: the US National Crime Survey, conducted from 1972 to 1975, included a module on the victimisation experience of respondents, which was placed after an attitudinal module about crime for a random half of the sample. For this subsample, victimisation reports increased significantly compared to when the attitude module was omitted (Gibson et al., 1978; Cowan, Murphy and Wiener, 1978). A plausible explanation is that the attitude items stimulated memory for and willingness to report on victimisation experiences (which are usually considered to be factual).

This highlights two potentially important messages for trust measures: first, different types of trust measures, ranging from evaluative to experience-based, can be prone to order effects. Placing them as early as possible in the survey should mitigate interference from other questions. It should be noted that as early as possible does not imply at the very beginning, without giving the interviewer the chance to build some degree of rapport with the respondent. Overall, it will be most important to avoid asking trust questions immediately after items that are likely to elicit strong emotional responses or that refer to experiences with other people or institutions. Second, it will be essential to consider not only the priming effect on the trust items in question, but also the priming effect these questions themselves can have on subsequent items, especially if they deal with similar content.

Priming can also occur with regard to broader environmental effects other than those concerning the immediate survey context. For example, in the case of subjective well-being measures, Deaton (2011) found that impactful short-term events, such as major news events (in his case, the 2008 financial crisis) and seasonal holidays, were associated with specific bumps in US time series data. There could be good reasons to assume that such events also impact mood and responses to interpersonal and institutional trust questions. Although financial crises and events such as terrorist attacks typically occur unexpectedly, regular events such as holidays, religious festivities and elections can be taken into account in survey scheduling and should probably be avoided. As with subjective well-being measures, it would be preferable to stage data collection throughout the year or at least over multiple days and weeks to minimise the influence of external events on responses (OECD, 2013).

Key messages

- Although order effects do not appear in every case and every survey, they can have a significant impact on responses when they do and should not be dismissed lightly.
- Order effects occur most often when two or more questions deal with the same or closely related issues, and initial evidence backs a mitigation strategy that either separates trust items within the survey as much as possible without destroying the coherence of the questionnaire or uses intervening text as a buffer.
- Whenever lists of trust items are used, two rules apply: first, general summary-type questions that ask respondents to evaluate complex topics in an overall way seem to be more sensitive to position than are more specific questions. Therefore, a survey should move from a broad to a narrow level of specificity within a group of questions, e.g. by placing items about generalised trust before questions about limited trust. Second, in order to control for salience effects (where the first item in a list is more likely to be

affected by order effects), which is especially important when using a list of different institutions, the order of items should be randomised.

- Trust measures should be placed early enough in the survey to avoid interference from other questions, but late enough to allow for bonding between interviewer and respondent. Overall, it will be key to avoid asking trust questions immediately after items that are likely to elicit strong emotional responses or that refer to experiences with other people or institutions. Questionnaire designers should also not forget about the potential effect that trust questions themselves can have on subsequent items, in particular those dealing with similar content.
- In order to minimise the impact of the broader external survey context, including holidays, seasons and elections, it would be preferable to stage data collection throughout the year or at least over multiple days and weeks.

3.6. Survey mode

Surveys can be conducted in a variety of ways. These include self-administered questionnaires (SAQs), traditionally conducted in a pen-and-paper format, but which increasingly involve internet-online surveys; computer-assisted self-interviews (CASI); telephone interviews and computer-assisted telephone interviews (CATI); and pen-and-paper interviewing (PAPI) and computer-assisted personal interviews (CAPI), usually conducted through visits to the survey respondent's home (OECD, 2013). Recent years have also seen a rise in mixed-method data collection, combining several of the above-listed modes in the same survey. The main distinction between the different survey modes is usually drawn between self-administered or interviewer-led surveys (Holbrook et al., 2003; Metzger et al., 2000; Turner et al., 2005; Tourangeau and Yan, 2007). Different survey modes can substantially influence how respondents process and reply to questions, as well as how much information they feel comfortable to reveal.

The issues

In practice, the choice of survey mode will be influenced by a variety of factors, including coverage and availability of sample frames, financial costs and fieldwork time, as well as the suitability of the questionnaire (Roberts, 2007). On top of that, the potential for error caused by survey mode should also feature in the survey mode selection process for trust questions. Coverage error, non-response error, measurement error and processing error can all be influenced by survey mode (ibid.).

This section focuses on the potential for measurement error, namely satisficing, or the use of response biases and heuristics by respondents, and sensitivity. All survey modes vary substantially in the pace with which the survey is conducted, the extent to which the flow of questions is determined by the interviewer or the respondent, whether a respondent can revisit questions throughout the survey, which types of visual aids are presented to the respondent, the extent of human interaction between interviewer and respondent, as well as the privacy of answers. These dimensions can substantially impact how respondents understand questions and portray themselves.

There are theoretical reasons to assume that there is a higher risk of satisficing in self-administered vs. interviewer-led modes (see Roberts, 2007). In self-administered modes there is no time pressure but also no interviewer to facilitate or motivate. Especially in

online surveys, multitasking is possible. Therefore, the overall cognitive burden and risk of satisficing in self-administered surveys is higher.

Since the concept of sensitivity has not been introduced yet in this chapter, a few explanatory words are warranted. The mode, especially through the degree of privacy that it affords, might cause certain types of self-presentational behaviour by respondents, sometimes with and sometimes without their own knowledge. For example, respondents might want to appear consistent across answers (e.g. if a respondent answers first that he or she approves of the current political leadership, he or she might also be tempted to indicate being satisfied with democracy and trusting in public institutions later on so as not to appear erratic). Social desirability is another example of self-presentational behaviour and relates to the difficulty of truthfully reporting an attitude or behaviour that violates existing social norms and may be deemed inappropriate by society. To conform to social norms, respondents may present themselves in a positive light, independently of their actual attitudes and true behaviours. More specifically, respondents might tend to admit to socially desirable traits and behaviours and to deny socially undesirable ones. This is an issue for data quality, as in the case of socially undesirable activities, sample proportions will underestimate the true prevalence and frequency of the attitude while simultaneously overestimating the true level of socially desirable behaviour. This might be an issue, specifically for questions on trust, if it is societally frowned upon to openly declare distrust of specific population groups, or if it is the prevailing fashionable norm to be sceptical of public institutions.

Social desirability is a distinct aspect of what can be termed the "sensitivity" of questions. While the issue associated with social desirability is the sensitivity of an answer, the sensitivity of the question itself poses a challenge if the question topic is deemed to be intrusive (too private or taboo in the culture of the respondent) or if it would be risky to the respondent if his or her true answers were made public or known to third persons beyond the survey setting. The costs and negative consequences could include prosecution or job loss. The reason for their sensitivity is likely to be different for interpersonal and institutional trust: it is plausible to assume that while items involving interpersonal trust are more likely to be affected by social desirability bias (it might not be considered socially acceptable to not trust other people, in particular members of another religion or nationality), while threat of disclosure might be an issue especially for questions about trust in institutions, particularly if the item is included in an official survey conducted by the government or its statistical agency. In terms of the effect on data quality, questions that are too intrusive or pose a threat if disclosed can increase the unwillingness to reply or result in a large number of missing values and *don't know* responses or in an overestimation of true attitudes if respondents fall back on the "publicly safe" answer.

A common approach to assess whether certain questions are prone to sensitivity-related response biases is a post-hoc assessment via empirical indicators of survey quality such as item non-response rates (Lensvelt-Mulders, 2008). Chapter 2 of these Guidelines has already analysed the item-specific non-response rates for trust questions for the Gallup World Poll and the European Social Survey and highlighted that, while all trust questions perform better than income questions, the bulk of trust questions have higher item-specific non-response rates than more straightforward questions on marital status, education or gender. Institutional trust questions perform worse than interpersonal trust questions in this analysis. Even religion, which is often considered a sensitive topic to ask about, had a non-response rate of less than half of most institutional trust questions. This suggests that trust questions, in particular those concerning institutional trust, can indeed be considered sensitive.

The evidence

While there are very few experimental studies specifically looking at trust and measurement error in relation to survey mode, many lessons from the literature on other self-reported items are applicable here.

Considering the impact of survey mode on the use of response biases and heuristics, there is indeed some evidence supporting the theory that satisficing in its various forms is more likely in self-administered modes than in interviewer-administered modes. For example, higher levels of acquiescence were found in a mixed-mode design that included self-administration than in one which mixed interviewer-administered modes only (Beukenhorst and Wetzels, 2009), and several scholars reported higher levels of *don't know* responses by self-administered (internet) surveys than in interviewer-led surveys (telephone, face-to-face) (see Duffy et al., 2005; Heerwegh, 2009). However, the evidence for satisficing in self-administered modes is not overwhelming – many times, the reported differences were either not significant or significant at very low levels.

Regarding the issue of sensitivity, the tables turn and self-administered survey modes perform much better compared to interviewer-led techniques. Various experimental field studies have established strong evidence that self-administered survey modes, in comparison to interviewer-led techniques, increase the levels of reporting socially stigmatised medical conditions such as depression or sexually transmitted diseases (Villarroel et al., 2008; Krumpal, 2013), socially undesirable activities such as illicit drug use, risky sexual behaviour and abortions (Gribble et al., 1999; Tourangeau and Yan, 2007), as well as socially unaccepted attitudes about race and same-gender sex (Krysan, 1998; Villarroel et al., 2006). Trust-specific evidence pointing in the same direction comes from Cycle 27 of the Canadian General Social Survey on Social Identity, which included questions on interpersonal and institutional trust. Prior to Cycle 27, data collection was done using only computer assisted telephone interviews (CATI); Cycle 27 was the first cycle where an internet self-response option was offered to respondents, and close to 7 000 out of roughly 20 000 respondents completed the survey using the internet option. Using propensity score matching to control for other mode effects (e.g. non-response bias, selection bias and process bias), Statistics Canada compared the responses of both types of data collection and found that CATI respondents showed significantly higher trust scores than did internet respondents.

Thus, while self-administered surveys, in particular web-based ones, carry a slightly higher risk of satisficing, there is a strong case to place trust items in self-administered modules whenever possible, for example in increasingly common mixed-method data collection procedures of national statistical offices (NSOs).

If a face-to-face survey is the only option for data collection, a couple of rules can mitigate the impact of sensitivity-related biases. First, there are documented effects of interviewers' characteristics (e.g. gender and socio-economic status) and assumed interviewer expectations on social desirability bias: Katz (1942) found increased reporting of pro-labour attitudes when interviews were conducted by working-class interviewers. Enumerators should therefore reveal as little of their own social identity as possible during interviews.

Second, several innovative interview methods can be applied to enhance respondents' feeling of privacy: for example, the "sealed envelope technique" (De Leeuw, 2001; Bradburn and Sudman, 1979) involves handing a separate self-administered questionnaire to the respondent in the sensitive questions part of the interview. Respondents are then asked to

complete the questionnaire, place it in an envelope, seal it and return it to the interviewer. Another method, the "unmatched count technique", involves randomly splitting the sample into two (Biemer and Brown, 2005). One group of respondents is asked to answer a short list of questions that includes only a set of non-sensitive items. The other subsample has to respond to a longer list consisting of the same non-sensitive items plus sensitive ones. Without telling the interviewer which specific items were answered *yes*, respondents in both groups count the number of positive answers and report solely the sum of these items. An unbiased estimate of the population's proportion not trusting specific groups or institutions can be obtained by calculating the difference between the two subsample means. While the unmatched count technique is quite innovative and has been successfully applied across a range of stigmatised behaviours (e.g. Haushofer and Shapiro, 2016), it can deal only with *yes/no* binary response formats at the moment and does not allow for individual-level analysis.

A third way to heighten respondents' sense of privacy and willingness to co-operate, both in face-to-face and self-administered survey situations, is to highlight confidentiality and data protection assurances at the beginning of the survey. Singer et al. (1995) reviewed the experimental literature on the effects of confidentiality assurances in questionnaire introductions. Although the average effect size was small, the authors found that such confidentiality assurances resulted in lower item non-response and higher response accuracy for sensitive items (including income). While many NSOs might already make use of data protection assurances, it is worth emphasising these even more when trust and other sensitive questions are included in the questionnaire.

Key messages

- While there is some evidence that self-administered surveys carry a higher risk of satisficing, this evidence is neither consistent nor overwhelming.
- There are ways of appropriately tailoring the survey design to reduce social desirability bias and concerns about the threat of disclosure. Sensitivity-related response biases can be reduced by increasing the anonymity of the question-and-answer process (e.g. through self-administered interviews), by decreasing the respondent's concerns about data protection (e.g. via confidentiality assurances), or by controlling the survey situation (e.g. not having enumerators give out information about their own social identity).
- While placing trust questions in self-administered surveys is strongly preferred, the use of innovative interviewing methods such as the sealed envelope or unmatched count technique could be explored in face-to-face surveys.

3.7. Response styles and the cultural context

The previous sections of this chapter have reviewed various methodological features that can affect response biases and data quality. In addition, individual respondents themselves can consistently be prone to certain forms of response biases or repeatedly rely on particular response heuristics. This type of constant response pattern is known as response style.

The issues

If a respondent consistently relies on a specific style of answering questions, a systematic bias across self-reported variables can be generated. This noise can translate into artificially higher correlations between these self-reported measures, an issue referred

to as common method variance (OECD, 2013). Since all trust measures are self-reported, they are potentially affected by response styles and common method variance.

Respondents are said to be particularly likely to rely on response styles as default patterns of answering if they are fatigued or confused by the way a question is presented or because of lack of knowledge or memory failure. It has been suggested that certain groups of respondents may be more likely to rely on response styles than others, for example people with lower cognitive skills (Krosnick, 1999). Another sometimes-cited factor behind response styles is a respondent's temperament and character (Spector et al., 2000), which can tip a response pattern towards more negative or more positive answers, depending on whether a person is more optimistic or pessimistic.

Beyond individual variation in response styles, a particular concern for international comparisons is the extent to which respondents from different cultures or linguistic groups might exhibit different response styles when answering trust and other self-reported questions. If it can be demonstrated that different countries systematically rely on different response style patterns, including scale use, the accuracy of comparisons between them may be limited.

The evidence

The evidence on response styles is not specific to trust items, but findings from other self-reported measures, including subjective well-being, are applicable and will be drawn upon in the following.

The literature on response styles and subjective measures paints a mixed picture, with some reviews finding a significant influence of acquiescence and other systematic errors on affect measures (see Watson and Clark, 1997). However, a number of other studies suggest that response styles have only a negligible effect on the level of subjective well-being measures (see Moum, 1988; Schimmack, Böckenholt and Reisenzein, 2002). It is also debated whether differences in response styles between population subgroups impact the overall validity of results in practice. For example, some studies have found that "nay-saying" is more common among younger people than older people, as well as among respondents from higher educational backgrounds (e.g. Gove and Geerken, 1977). Importantly, though, these differences did not alter overall relationships between socio-demographic variables (including income, occupation, marital status, race, gender, age and education) and self-reported mental well-being (OECD, 2013).

In general, the biggest challenge to identifying whether a response style is present and to quantifying its impact on results is that response patterns are extremely difficult to verify externally against a common standard or actual behaviour. Usually, response styles are "detected" by comparing a respondent's choices across a variety of survey items: if someone selects the (positive or negative) extremes of scales consistently, or chooses to agree systematically with self-reported statements, he or she is considered to follow an extreme responding or acquiescent response style. For example, Marín, Gamba and Marín (1992) estimate acquiescence through counting the number of times a respondent agreed with a question and then created an extreme responding indicator by counting the times a respondent selected either of the scale anchors. However, unless the responses are logically contradictory (e.g. agreeing to a statement that no one can be trusted and to a statement that everyone can be trusted), it is difficult to tell whether these responses are due to consistent response biases or indeed genuinely reflect the respondent's feelings and level of trust. Thus

we often do not know whether a pattern adds error to the data or represents meaningful variation in trust.

Studies that examine differences in response styles between countries run into similar problems. On the one hand, a couple of works suggest that response styles do indeed vary between cultures. For example, Van Herk, Poortinga and Verhallen (2004) examined marketing data from six EU countries and discovered systematic differences of medium effect size in acquiescence and extreme response styles, with both styles being more prevalent in data from Mediterranean countries (Greece, Italy and Spain) than from northwestern Europe (Germany, France and the United Kingdom). Marín, Gamba and Marín (1992), Clarke (2001) and Holbrook et al. (2006) all suggest that US Hispanics and African-Americans prefer more extreme response categories and are more likely to acquiesce compared to samples of US whites. In contrast, Asian Confucian cultures have been found to be more likely to respond moderately (Lau, Cummins and McPherson, 2005; Lee, et al., 2002) and to be more prone to social desirability bias (Abe and Zane, 1990; Middleton and Jones, 2000) than less collectivist Western nations. However, with the exception of the 2004 Van Herk, Poortinga and Verhallen paper, which includes measures of actual behaviour, all other study designs do not include such validation and therefore do not provide conclusive evidence that the response patterns add noise to the results. Study participants may very well have selected more or less extreme responses and agreed more or less with statements because they represented how they actually feel, rather than how they respond to questions. Furthermore, a recent cross-country study by Exton, Smith and Vandendriessche (2015) drawing on multiple Gallup World Poll waves concluded that culture (including measurement error and actual differences in the experience of life) may account for at most 20% of unexplained country-specific variance in subjective well-being. This effect is small when compared to the role of objective life circumstances in explaining subjective well-being outcomes.

Therefore, both for single and multi-country studies, it is safe to conclude that even in the cases where it can be proven that a response pattern is unambiguously due to differences in response styles and not to actual respondent evaluations, response styles do not seem to harm overall data quality to such a degree that trust and other self-reported measures should be dismissed as invalid. This insight also has implications for the appropriate strategies to deal with response styles: several scholars have suggested either controlling for the factors assumed to drive response styles (e.g. personality) in the analysis or going a step further and applying statistical adjustment techniques such as mean correction or scale standardisation directly to measures believed to be affected by response styles (Greenleaf, 1992; Hofstede, 2001). However, there is a substantial risk that such strategies would eliminate true substantive differences (see Harzing, 2006). Rather than trying to eliminate response bias retrospectively through statistical adjustment, it might therefore be preferable to avoid response style bias in the first place by careful questionnaire design. For example, questions with *agree/disagree* and to a lesser degree *yes/no* response formats might be more likely to prompt acquiescence and should be avoided if possible (Krosnick, 1999). Smith (2003) also suggests that the use of a mixture of positive and negative statements will mitigate both acquiescence and disacquiescence, because it might lead respondents to consider the exact meaning of the question more closely and as a result give more meaningful responses, or at least lead to the responses cancelling each other out. This approach will need to be further tested, as it poses a risk of confusing respondents when the same scale end presents something positive in one item and something negative in a following one. Moreover, questionnaire items containing negations can be difficult to translate into some languages.

In general, given that individuals are assumed to be more likely to rely on response biases and heuristics when they are confused by questions, less motivated, more fatigued and more burdened, the best way to minimise these issues is likely to be through adopting solid survey design principles: avoiding items that are difficult to understand or repetitive or that look too similar; using short and engaging questions that are easy to answer; and keeping respondents interested and motivated (OECD, 2013). These principles are true for all survey measures, and there is no strong reason to assume that trust measures are at any greater risk of eliciting response styles than other self-reported survey items.

Key messages

- Response styles are very difficult to verify externally against a common standard or actual behaviour. More often than not, we do not know whether a pattern adds error to the data or represents meaningful variation in trust.
- Even where the existence of response styles has been established, they do not necessarily seem to harm overall data quality to such a degree that trust and other self-reported measures should be dismissed as invalid.
- Rather than using statistical adjustment techniques to mitigate response style bias, the focus should be on designing the questionnaire so that items are as simple, easy to interpret and minimally burdensome as possible. The overall survey design (including its length and how it is introduced) needs to pay particular attention to respondent burden, motivation and fatigue in order to maximise data quality.
- Question formats that are known to be more prone to response biases should be avoided. A case in point are *agree/disagree* and to a lesser degree *yes/no* response formats, which are more likely to prompt acquiescence.
- For internationally comparative analyses of trust data, one option to get around response style concerns could be to use changes in response patterns over time (including those of different population subgroups) rather than the level of responding.

3.8. Conclusion

This chapter has discussed the various methodological features of question and survey design that can impact measurement error for trust items. Key conclusions and areas warranting future research include the following:

- While trust measures are more sensitive to response biases than more objective measures (such as educational attainment or life expectancy), these biases are also likely to occur in other self-reported measures that are already being collected by NSOs. Although it is essential to be aware of these biases and of the most appropriate question and survey design strategies to mitigate them, the existence of measurement error *per se* is not an argument against gathering data on trust. Especially for items on interpersonal trust, the evidence of their strong validity with regard to real world outcomes demonstrates that these measures are meaningful and worth collecting.
- No matter which approach to question design is adopted by data collectors, standardisation is critical to ensure meaningful comparison over time and between groups and countries.
- The evidence on question wording (especially that drawn from split sample experiments) shows that this is not a trivial matter and that good question wording matters for results. Question wording should avoid referring to concepts other than trust and be specific and precise to the situation of interest. For interpersonal trust, a neutral question wording is

recommended: data collectors should refrain from referring to *caution in dealing with other people*, as this wording can prime more vulnerable groups to report lower trust. For institutional trust, specifying what institutions are expected to do can make a difference in some cases. Overall, question wording should be precise enough to be understood by respondents, without getting into subtle nuances (which might also pose an issue for translatability across countries). If the concepts that different questions try to capture are too narrowly related, respondents might have difficulty differentiating between them (e.g. trust vs. confidence).

- The way answering options are presented can have a significant impact on the distribution of responses. For trust items, a numerical 0-10 scale with verbal scale anchors is recommended, as it allows for a high degree of variance in responses, increases overall data quality and facilitates translatability across languages. The response order should be presented consistently (i.e. 0-10 instead of 10-0) in order to minimise mental switching between positive and negative normative outcomes. The verbal descriptions of the scale anchors should represent absolute responses (e.g. *completely/not at all*) to minimise acquiescence bias and socially desirable responding and to allow for the full spectrum of possible responses.
- Trust measures should be considered within the broader survey context in which they are placed. As with the standardisation of wording and response formats, consistency of order within question modules across surveys and over time is essential to guarantee the quality and comparability of trust measures. Since order effects occur most often when two or more questions deal with the same or closely related issues, trust items should either be separated within the survey as much as possible or buffered by intervening text. Whenever lists of trust items are used, a survey should move from a broad to a narrow level of specificity within a group of questions, e.g. by placing items about generalised trust before questions about limited trust. When thinking about placement of trust questions in the survey, a balance is needed between showing questions early enough to avoid interference from other questions and late enough to allow for bonding between interviewer and respondent. Generally, trust questions should not be asked immediately after items that are likely to elicit strong emotional responses or that refer to experiences with other people or institutions. Questionnaire designers should equally reflect on the potential effect that trust questions themselves can have on subsequent items, in particular those dealing with similar content. Lastly, in order to minimise the impact of holidays, seasons and elections, data collection is recommended to be spread throughout the year or at least over multiple weeks.
- Evidence suggests that trust questions can be sensitive, triggering respondents to answer in socially desirable way or be unwilling to answer at all. This is especially true for measures of trust in institutions. Self-administered surveys, compared to interviewer-led ones, perform better in terms of minimising social desirability. This benefit outweighs the (relatively weak) counterargument that self-administered surveys carry a higher risk of satisficing. In all survey modes, sensitivity-related response biases can be reduced by decreasing the respondent's concerns about data protection (e.g. via confidentiality assurances) or by controlling the survey situation (e.g. not having enumerators give out information about their own social identity). If face-to-face interviews are the only option, the use of innovative interviewing methods such as the sealed envelope or unmatched count technique could be explored.

- Cross-cultural response styles are very difficult to verify externally against a common standard or actual behaviour. Even where the existence of response styles has been established, they do not necessarily harm overall data quality to such a degree that trust and other self-reported measures should be considered as invalid. If data producers want to mitigate the possibility of response style bias, they should, rather than relying on *ex post* statistical adjustment techniques, focus on designing the questionnaire so that items are as simple, easy to interpret and minimally burdensome as possible. The overall survey design (including its length and how it is introduced) needs to pay particular attention to respondent burden, motivation and fatigue in order to maximise data quality. Moreover, question formats that are more prone to response biases should be avoided: for example, *agree/disagree* and to a lesser degree *yes/no* response formats are more likely to prompt acquiescence.
- Further research is needed on both institutional and interpersonal trust, but especially on the former, for which there is very little methodological evidence available.
  - First, with regard to question wording for institutional trust, experimental testing should be used to establish which specifications other than *to act in the national interest* (e.g. *to improve the life of someone like me* or *to do what is right*) matter for which institutions. Ideally, these experiments should be carried out across more than just one country.
  - Second, while it has been suggested that the use of a mixture of positive and negative statements can mitigate both *yay* and *nay* saying, this approach needs to be further tested to rule out the risk of confusing respondents when the same scale end presents something positive in one item and something negative in a following one.
  - Third, with regard to order effects, it is not yet clear in which cases these occur for trust questions. More targeted methodological research is needed to discover what type of question or what type of context triggers which effect in order to further inform survey design. While there is some evidence that transitional text between questions can act as a buffer to mitigate order effects, various text versions should be tested for their impact on trust questions.
  - Finally, more research that validates response styles from different cultures against external references, such as actual trusting behaviour in real life or experimental games, would enrich the current body of cross-cultural trust research.

Notes

1. The four interpersonal trust question versions in the ESC survey were: 1) "Generally speaking, would you say that most people can be trusted or that you can't be too careful in dealing with people?"; 2) "People can be trusted until they prove otherwise"; 3) "Generally speaking, most people can be trusted. Do you agree or disagree?" and 4) "Generally speaking, you can't be too careful in dealing with people. Do you agree or disagree?"
2. Evidence of perceived vulnerability of these population groups can be found in responses on whether they feel safe walking alone at night in their neighbourhood, which demonstrate a significantly lower proportion of women and people over 50 years reporting to feel safe compared to other population groups (OECD, 2015).
3. The OECD also partnered with INEGI of Mexico, placing various trust question versions in Mexico's June 2016 National Survey on Urban Public Security. This study was not a split sample experiment (each respondent was asked two questions within the same survey), and therefore it cannot be ruled out that differences in responses are due to priming or shared method variance. Nevertheless, when the two versions (regular trust in institutions questions vs. *to act in the national interest*) were posed to the 500-person sample, adding *to act in the national interest* did not lead to a strong drop in the

population share which indicated that banks can be trusted *a great deal*. On the other hand, the civil service experienced a drop from 10.6% to 6.8% of respondents indicating that they trust this institution *a great deal*. These results potentially indicate that institutions could carry different connotations in the context of Mexico compared to the UK. Further actual experimental research will be needed to clarify this issue.

4. While some NSOs, such as the Central Statistical Office of Poland, have introduced verbally labelled 4-point scales, comparative methodological evidence on the effect of this type of response scale remains limited.

References

Abe, J.S. and N.W.S. Zane (1990), "Psychological maladjustment among Asian and White American college students: Controlling for confounds", *Journal of Counselling Psychology,* Vol. 37, pp. 437-444.

Alwin, D.F. and J.A. Krosnick (1991), "The reliability of survey attitude measurement: The influence of question and respondent attributes", *Sociological Methods and Research*, Vol. 20, No. 1, pp. 139-181.

Beukenhorst, D. and W. Wetzels (2009), "A comparison of two mixed mode designs of the Dutch Safety Monitor: Mode effects, costs, logistics", Paper presented at the European Survey Research Association Conference, Warsaw.

Biemer, P.P. and G. Brown (2005), "Model-based estimation of drug use prevalence using item count data", *Journal of Official Statistics*, Vol. 21, No. 2, pp. 287-308.

Bradburn, N.M. and S. Sudman (1979), *Improving Interview Method and Questionnaire Design*, Jossey-Bass, San Francisco.

Bradburn, N.M., S. Sudman and B. Wansink (2004), *Asking Questions: The Definitive Guide to Questionnaire Design – from Market Research, Political Polls, and Social and Health Questionnaires*, Jossey-Bass, San Francisco.

Chang, L. (1994), "A psychometric evaluation of 4-point and 6-point Likert-type scales in relation to reliability and validity", *Applied Psychological Measurement*, Vol. 18, pp. 205-215.

Clarke, J. (2001), "Extreme response style in cross-cultural research", *International Marketing Review,* Vol. 18, pp. 301-324.

Cowan, C.D., L.R. Murphy and J. Wiener (1978), "Effects of supplemental questions on victimization estimates from the national crime survey", *1978 Proceedings of the Section on Survey Research Methods,* American Statistical Association, Washington, DC, United States.

Cummins, R. and E. Gullone (2000), "Why we should not use 5-point Likert scales: The case for subjective quality of life measurement", *Proceedings of the Second International Conference on Quality of Life in Cities,* Singapore.

Deaton, A.S. (2011), "The financial crisis and the well-being of Americans", *Working Paper*, No. 17128, National Bureau of Economic Research (NBER), Cambridge MA, *www.nber.org/papers/w17128*.

De Leeuw, E.D. (2001), "Reducing missing data in surveys: An overview of methods", *Qual. Quant.*, Vol. 35, pp. 147-160.

Denters, B., O. Gabriel and M. Torcal (2007), "Political confidence in representative democracies: Social capital vs. political explanations", J. van Deth, J.R. Montero and A. Westholm (eds), *Citizenship and Involvement in European Democracies*, Routledge, Abington.

Duffy, B. et al. (2005), "Comparing data from online and face-to-face surveys", *International Journal of Market Research*, Vol. 47, No. 6.

Exton, C., C. Smith and D. Vandendriessche (2015), "Comparing happiness across the world: Does culture matter?", *OECD Statistics Working Papers*, No. 2015/04, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jrqppzd9bs2-en*.

Gibson, C. et al. (1978), "Interaction of survey questions as it relates to interviewer-respondent bias", *1978 Proceedings of the Section on Survey Research Methods,* American Statistical Association, Washington, DC, United States.

Gove, W.R. and M.R. Geerken (1977), "Response bias in surveys of mental health: An empirical investigation", *American Journal of Sociology,* Vol. 82, No. 6, pp. 1289-1317.

Greenleaf, E.A. (1992), "Improving rating scale measures by detecting and correcting bias components in some response styles", *Journal of Marketing Research,* Vol. 29, pp. 176-188.

Gribble, J.N. et al. (1999), "Interview mode and measurement of sexual behaviors: Methodological issues", *Journal of Sexual Research*, Vol. 36, pp. 16-24.

Harzing, A.W. (2006), "Response styles in cross-national survey research: A 26-country study", *International Journal of Cross-Cultural Management,* Vol. 6, pp. 243-266.

Haushofer, J. and J. Shapiro (2016), "The short-term impact of unconditional cash transfers to the poor: Evidence from Kenya", *Quarterly Journal of Economics*, Vol. 131, No. 4, pp. 1973-2042.

Heerwegh, D. (2009), "Mode differences between face-to-face and web surveys: An experimental investigation of data quality and social desirability effects", *International Journal of Public Opinion Research*, Vol. 21, No. 1, pp. 111-121.

Hofstede, G.H. (2001), *Culture's Consequences: Comparing Values, Behaviours, Institutions, and Organizations across Nations*, Sage, Thousand Oaks, California.

Holbrook, A.L., T.P. Johnston and Y.I. Cho (2006), "Extreme response style: Style or substance?", Paper presented at the 61st Annual Meeting of the American Association for Public Opinion Research, Montreal, Canada.

Holbrook, A.L., M.C. Green and J.A. Krosnick (2003), "Telephone versus face-to-face interviewing of national probability samples with long questionnaires: Comparisons of respondent satisficing and social desirability response bias", *Public Opinion Quarterly*, Vol. 67, pp. 79-125.

Hooghe, M. and T. Reeskens (2008), "Cross-cultural measurement equivalence of generalized trust: Evidence from the European Social Survey (2002 and 2004)", *Social Indicators Research,* Vol. 85, pp. 15-32.

Jagodzinski, W. and K. Manabe (2004), "How to measure interpersonal trust? A comparison of two different measures", *ZA-Information / Zentralarchiv für Empirische Sozialforschung,* Vol. 55, pp. 85-98.

Katz, D. (1942), "Do interviewers bias poll results?", *Public Opinion Quarterly*, Vol. 6, pp. 248-268.

Krosnick, J.A. (1999), "Survey research", *Annual Review of Psychology*, Vol. 50, pp. 537-567.

Krosnick, J.A. (1991), "Response strategies for coping with the cognitive demands of attitude measures in surveys", *Applied Cognitive Psychology,* Vol. 5, pp. 213-236.

Krumpal, I. (2013), "Determinant of social desirability bias in sensitive surveys: A literature review", *Qual Quant*, Vol. 47, pp. 2025-2047.

Krysan, M. (1998), "Privacy and the expression of white racial attitudes – a comparison across three contexts", *Public Opinion Quarterly*, Vol. 62, pp. 506-544.

Lau, A.L.D., R.A. Cummins and W. McPherson (2005), "An investigation into the cross-cultural equivalence of the personal wellbeing index", *Social Indicators Research*, Vol. 72, pp. 403-430.

Lee, J.W. et al. (2002), "Cultural differences in responses to a Likert scale", *Research in Nursing and Health*, Vol. 25, pp. 295-306.

Lensvelt-Mulders, G.J.L.M. (2008), "Surveying sensitive topics", De Leeuw, E.D., J.J. Hox, D.A. Dillman (eds), *The International Handbook of Survey Methodology,* Erlbaum/Taylor&Francis, New York/London.

Lundasen, S. (2010), "Methodological problems with surveying trust", manuscript.

Maggino, F. (2009), "Methodological aspects and technical approaches in measuring subjective well-being", Università degli Studi di Firenze, Working Paper.

Marín, G., R.J. Gamba and B.V. Marín (1992), "Extreme response style and acquiescence among Hispanics: The role of acculturation and education", *Journal of Cross-Cultural Psychology*, Vol. 23, Vol. 4, pp. 498-509.

Metzger, D.S. et al. (2000), "Randomized controlled trial of audio computer-assisted self-interviewing: Utility and acceptability in longitudinal studies. HIVNET Vaccine Preparedness Study Protocol Team", *American Journal of Epidemiology*, Vol. 152, Vol. 2, pp. 99-106.

Middleton, K.L. and J.L. Jones (2000), "Socially desirable response sets: The impact of country culture", *Psychology and Marketing*, Vol. 17, No. 2, pp. 149-163.

Moum, T. (1988), "Yea-saying and mood-of-the-day effects in self-reported quality of life", *Social Indicators Research*, Vol. 20, pp. 117-139.

Newstead, S.E. and J. Arnold (1989), "The effect of response format on ratings of teaching", *Educational and Psychological Measurement*, Vol. 49, pp. 33-43.

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

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, Oct 2015 to May 2016", released 10 November 2016, Office for National Statistics, UK, ONS, Newport, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

Peadody, D. (1961), "Authoritarianism scales and response bias", *Psychological Bulletin*, Vol. 65, No. 1, pp. 11-23.

Podsakoff, P.M. et al. (2003), "Common method biases in behavioral research: A critical review of the literature and recommended remedies", *Journal of Applied Psychology*, Vol. 88, No. 5, pp. 879-903.

Preston, C.C. and A.M. Colman (2000), "Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences", *Acta Psychologica*, Vol. 104, pp. 1-15.

Pudney, S. (2010), "An experimental analysis of the impact of survey design on measures and models of subjective well-being", *Institute for Social and Economic Research Working Papers,* No. 2010-20, University of Essex.

Roberts, C. (2007), "Mixing modes of data collection in surveys: A methodological review", *Economic and Social Research Council – National Centre for Research Methods*, NCRM Methods Review Papers NCRM/008.

Roberts, J. and M. Hough (2005), *Understanding Public Attitudes to Criminal Justice*, Open University Press, Maidenhead.

Saris, W.E. and I. Gallhofer (2006), "Report on the MTMM experiments in the pilot studies and proposals for Round 1 of the ESS", *www.europeansocialsurvey.org/docs/methodology/ESS1_quality_measurement.pdf*.

Scherpenzeel, A. and W.E. Saris (1997), "The validity and reliability of survey questions – A meta-analysis of MTMM studies", *Sociological Methods Research*, Vol. 25, No. 3, pp. 341-383.

Schimmack, U., U. Böckenholt and R. Reisenzein (2002), "Response styles in affect ratings: Making a mountain out of a molehill", *Journal of Personality Assessment*, Vol. 78, No. 3, pp. 461-483.

Schuman, H. and S. Presser (1996), *Questions and Answers in Attitude Surveys,* Sage Publications, California, United States.

Schwarz, N. and F. Strack (2003), "Reports of subjective well-being: Judgemental processes and their methodological implications", D. Kahneman, E. Diener and N. Schwarz (eds.), *Well-being: The Foundations of Hedonic Psychology*, Russell Sage Foundation, New York.

Schwarz, N. and H. Schuman (1997), "Political knowledge, attribution and inferred interest in politics", *International Journal of Public Opinion Research*, Vol. 9, No. 2, pp. 191-195.

Searle, J.R. (1979), *Expression and Meaning,* Cambridge University Press, New York, United States.

Singer, E., D.R. von Thurn and E.R. Miller (1995), "Confidentiality assurances and response: A quantitative review of the experimental literature", *Public Opinion Quarterly*, Vol. 59 No. 1, pp. 66-77.

Smith, T.W. (2003), "Developing comparable questions in cross-national surveys", J.A. Harkness, F.J. van de Vijver and P.P. Mohler (eds.), *Cross-Cultural Survey Methods*, Wiley, New York.

Smith, T.W. (1997), "Factors relating to misanthropy in contemporary American society", *Social Science Research,* Vol. 26, pp. 170-196.

Smith, T.W. (1981), "Can we have confidence in confidence? Revisited", Denis Johnston (eds.), *Measurement of Subjective Phenomena,* US Government Printing Office, Washington, DC.

Soroka, S., J. Helliwell and R. Johnston (2007), "Measuring and modelling trust", F. Kay and R. Johnston (eds.), *Diversity, Social Capital and the Welfare State,* University of British Colombia Press, Vancouver, BC.

Spector, P.E. et al. (2000), "Why negative affectivity should not be controlled in job stress research: Don't throw out the baby with the bath water", *Journal of Organizational Behavior*, Vol. 21, No. 1, pp. 79-95.

Stolle, D. (1998), "Bowling together, bowling alone: The development of generalized trust in voluntary associations", *Political Psychology,* Vol. 19, No. 3, pp. 497-525.

Sudman, S., N.M. Bradburn and N. Schwarz (1996), *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*, Jossey-Bass, San Francisco.

Tourangeau, R. and T. Yan (2007), "Sensitive questions in surveys", *Psychological Bulleting*, Vol. 133, pp. 859-883.

Turner, C.F. (2005), "Reducing bias in telephone survey estimates of the prevalence of drug use: A randomized trial of telephone audio-CASI", *Addiction* , Vol. 100, pp. 1432-1444.

Uslaner, E.M. (2009), "Is eleven really a lucky number? Measuring trust and the problem of clumping", Manuscript.

Uslaner, E.M. (2002), *The Moral Foundations of Trust,* Cambridge University Press, Cambridge, United Kingdom.

Van Herk, H., Y.H. Poortinga and T.M.M. Verhallen (2004), "Response styles in rating scales: Evidence of method bias in data from six EU countries", *Journal of Cross-Cultural Psychology*, Vol. 35, No. 3, pp. 346-360.

Villarroel, M.A. et al. (2008), "T-ACASI reduces bias in STD measurements: The national STD and behavior measurement experiment", *Sexually Transmittable Diseases*, Vol. 35, pp. 499-506.

Villarroel, M.A. et al. (2006), "Same-gender sex in the United States: Impact of T-ACASI on prevalence estimates", *Public Opinion Quarterly*, Vol. 70, pp. 166-196.

Watson, D. and L.A. Clark (1997), "Measurement and mismeasurement of mood: Recurrent and emergent issues", *Journal of Personality Assessment*, Vol. 68, No. 2, pp. 267-296.

Weng, L.J. (2004), "Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability", *Educational and Psychological Measurement*, Vol. 64, pp. 956-972.

Zmerli, S. and K. Newton, (2008), "Generalised trust and attitudes toward democracy", *Public Opinion Quarterly,* Vol. 72, pp. 706-724.

Zmerli, S., K. Newton and J.R. Montero (2007), "Trust in people, confidence in political institutions, and satisfaction with democracy", J. van Deth, J. R. Montero and A. Westholm (eds.), *Citizenship and Involvement in European Democracies*, Routledge, Abington.

© OECD 2017

Chapter 4

Measuring trust

This chapter provides concrete advice on best practice in measuring trust in household surveys. The chapter discusses how to plan for the measurement of trust and provides concrete advice on the survey and sample design, including the target population, sample size, frequency and the duration of enumeration. Following this, the chapter sets out specific advice on questionnaire design, identifying a set of core measures of trust that should be the highest priority for measurement and which represent a minimum viable set of measures. Finally, the chapter examines the issues involved when implementing surveys to measure trust, including data coding, as well as issues relevant to interview training.

4.1. Introduction

This chapter presents best practice in measuring trust. While it describes both interpersonal and institutional trust, much of the focus is on the former, as better evidence is available on its validity. Measurement of institutional trust is also discussed, but in the context of experimental measures that will assist in building a better understanding of how such measures perform. The chapter covers both the range of concepts to be measured and the best approaches for measuring them. This includes issues of sampling, survey design, data processing and coding, and questionnaire design. In particular, the chapter presents a single primary measure of generalised interpersonal trust intended to be collected consistently across countries, as well as a small group of core measures that data producers should collect where possible (Box 4.1). This core set includes measures of institutional trust as well as additional items related to interpersonal trust. Although the measures of institutional trust are more experimental than those for interpersonal trust, their high policy relevance warrants their inclusion in the core. Beyond this core suite of measures, the chapter provides more general advice to support data producers interested in identifying and measuring aspects of trust that will meet their particular research or policy needs, as well as a range of additional question modules covering different approaches to measuring trust.

Box 4.1. **Core measures of trust**

Core measures of trust are those for which international comparability is the highest priority. These are measures for which there is the most evidence of validity and relevance, for which the results are best understood, and for which the policy uses are most developed. Although the Guidelines are intended to support producers of measures of trust rather than being overly prescriptive, the core measures proposed here are quite specific in terms of content and collection method.

The core measures outlined in this chapter consist of five questions.

- The first is a primary measure intended to be collected consistently across countries. This measure should be regarded as the highest priority for national statistical agencies and focuses on generalised interpersonal trust. It should be the first question included in surveys where the measurement of trust is considered.
- The additional four questions in the core set aim to collect more information on limited interpersonal trust, and also include a three-item set of experimental questions on aspects of institutional trust. All these questions are important and should be collected where possible. However, it is recognised that not all national statistical offices will be able to collect these measures in their core surveys.

Beyond articulating a suite of core measures, the main goal of this chapter is to provide general advice to data providers. In particular, the chapter is intended to support national statistical agencies and other data providers in the process of deciding what to measure and how to implement the measurement process most effectively. While models are provided for specific questions, the chapter aims to provide options and advice rather than directions.

The chapter has four substantive sections. Section 4.2 focuses on issues associated with planning the measurement of trust, including the relationship between the intended policy or research use of the data and the appropriate measurement objectives; the range of covariates to be collected alongside trust to support analysis and interpretation is also discussed. Section 4.3 addresses survey and sample design issues; these include the choice of the survey vehicle, sample design, target population, collection period and survey frequency. Section 4.4 looks at questionnaire design, which includes question order, questionnaire structure and question wording: a key element of this section is the inclusion of model questions on interpersonal and institutional trust. Finally, section 4.5 focuses on survey implementation, which includes brief guidelines on interviewer training as well as data processing. Issues relating to the use and analysis of trust data are addressed in Chapter 5 (on the output and analysis of measures of trust).

4.2. What to measure? Planning the measurement of trust

This section looks at the planning stage of a measurement project. It is concerned with what concepts to measure and how these concepts affect decisions about the final output and analysis. Some of the issues touched on in this section are discussed in greater depth in Chapter 5 (on the output and analysis of trust measures). However, where Chapter 5 focuses on how to analyse, interpret and present trust data, the discussion here is limited to how users' needs determine what information to collect.

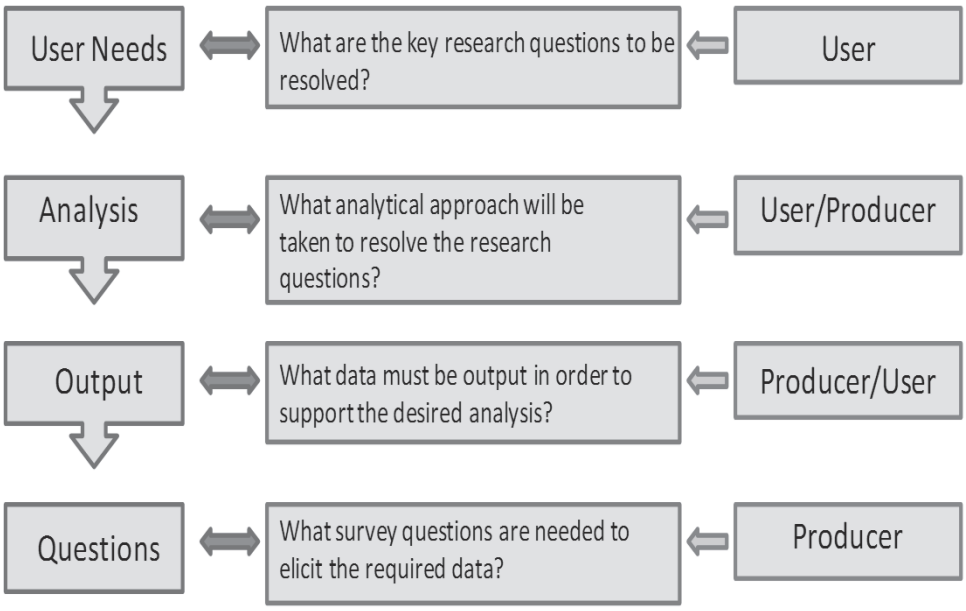
The initial planning stage of a project to measure trust – or indeed of any statistical programme – is critically important. All subsequent decisions are heavily influenced by choices made early on about the objectives of the project. Clarity about objectives is thus crucial.

Decisions about what to measure should always be grounded in a clear understanding of users' needs. Only if the needs of data users are clearly understood is it possible to make informed decisions about the information that should be collected to meet these needs. Understanding user needs is not, however, straightforward. A relatively simple research question can be approached in a range of different ways, using different methodologies. For example, one can understand what motivates behaviour both by asking people directly what they would do in a given set of circumstances or by collecting information on the course of action people take when facing a given set of circumstances. Each methodological approach has its own strengths and weaknesses and will have different implications for measurement. An analytical model can assist in thinking in a structured way about how user needs relate to specific decisions about what data to collect.

Figure 4.1 presents a simple model relating user needs to the specific survey questions used to collect information. The model is drawn from the *OECD Guidelines on Measuring Subjective Well-being* (2013), but is equally relevant to any survey design project; it provides a framework for thinking about the various stages involved in moving from a user's needs to specific questions to be included in a survey.

The first column of Figure 4.1 identifies the four stages involved in going from users' needs to specific survey content. Conceptually, these stages involve working back through the process of collecting the data and using them in decision making, in reverse order. Column 2 articulates the key issues to be addressed in each stage of the project in order to make well-informed decisions about the most appropriate measures. Finally, Column 3 indicates which party has the lead role in making decisions. Although the process of going

Figure 4.1. **The planning process: From user needs to survey questions**



Source: OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

StatLink http://dx.doi.org/10.1787/888933583823

from users' needs to survey content is fundamentally collaborative in nature, there are stages in the process when users can be expected to play a more important role than data providers, as well as stages where the reverse is true.

In practice, the process of working through these four stages is likely to be less clearly defined than Figure 4.1 suggests. In some cases, where the level of analysis required is relatively simple, the analysis and output stages of the process can merge into each other. Users will sometimes have clear views about the best measures to support the analysis that they would like to undertake, and it would be foolish to ignore these views in instances where a sophisticated user has a better understanding than a data provider of the issue at hand. Similarly, data producers may suggest possibilities that will result in changes in user needs or in the analytical approach taken to address them.

User needs

Understanding users' needs involves understanding the key policy and research questions that users are trying to address. While it is not possible in this chapter to give a full discussion of all possible users' needs for trust data, some general questions can be articulated:

- Are users' needs related to one of the general policy uses for trust data described in Chapter 2?
- What are the policy questions?
- Is the trust content being proposed appropriate to respond to these policy questions? Does the measure proposed allow monitoring changes over time or comparing population groups?

- What population groups are of greatest interest to the user? For example, is the focus on international comparisons (making countries the key unit of analysis), the same population at different points in time (time-series analysis), or different sub-groups of the same population (based on criteria such as age, gender, location or ethnicity)? These questions will have implications both for sampling and for the types of measure that are most appropriate. In the case of cross-country comparisons, measures with good cross-cultural reliability will be most important, while for analysis of groups within a country, low respondent burden may be a more important consideration in order to allow a larger sample size.
- Does the user's interest lie in comparing outcomes of different groups or in understanding the relationship between different aspects of trust? In the first case, a narrow range of trust measures may suffice, while in the latter case more detail on a range of covariates is likely to be necessary.
- Is the user's primary interest in generalised trust, in limited trust or in institutional trust? If the focus is on institutional trust, which institutions are of primary interest?
- What are the frequency requirements, i.e. the time periods over which the users need to monitor changes in trust?
- What within-country comparisons are required, such as geographic level?

A thorough understanding of users' needs should allow the identification of one or more clear research questions that the project should address.

Analysis

Understanding the overall research question is not sufficient to make meaningful decisions about the type of output or the most appropriate measures to use. A given research question may be addressed in more than one way. It is therefore essential to understand how the specific research question can be answered.

- Will the analytical approach be primarily descriptive, or will it require more sophisticated statistical techniques (e.g. regression, factor analysis, etc.)?
- What contextual and other variables are required to answer the research question? If the research question simply involves identifying differences between population groups in terms of a small set of key outcomes, the range of relevant covariates may be limited. However, if the research question is focused on understanding the drivers of group differences in levels of trust, or on examining the joint distribution of trust and other outcomes, the range of covariates is likely to be significantly broader.
- What level of accuracy is required to produce meaningful results from the proposed analysis? This will have implications for sample size and sampling strategy. For example, if obtaining precise estimates for small population sub-groups is a priority, then oversampling of these groups may be necessary.

After considering the proposed analytical strategy, it should be possible to articulate how the research questions can be answered in quite specific terms. This will form the basis for evaluating what data need to be produced to support the desired analysis.

Output

Output refers to the statistical measures released by a national statistical agency or by another data producer. These can take the form of tables of aggregate data (e.g. average

results by group), microdata files, interactive data cubes or other forms. The key distinction between output and analysis is that output does not, in itself, answer a research question. Rather, it provides the information base that is analysed in order to produce the answer. In some cases, the answer may be directly evident from the output, requiring only limited interpretation, comment and caveats, while in other cases extensive analysis may be required.

Because output forms the basis for all subsequent analysis, it provides the key link between specific survey questions and the use of data in the analysis. The required output must therefore be clearly specified before appropriate questions can be designed. Some key issues to consider when specifying the desired output for information on trust include:

- Will the analysis require tabular output of averages or proportions, or is microdata needed? Simple comparisons of how different population groups compare with each other can be accomplished via tabular output, but understanding the drivers of such differences will require a much finer level of detail.
- Will the analytic techniques used treat the data as ordinal or cardinal? This makes little difference if microdata is required (since users can decide for themselves), but it will influence how summary measures of central tendency and distribution are presented in tabular form. Information on a cardinal variable[1] can be presented via techniques that add and average scores (e.g. mean, standard deviation), while ordinal data will need to be reported by category.
- How important is it to present measures of the central tendency of the data (e.g. mean, median, mode) as opposed to the dispersion (e.g. standard deviation) or to the full distribution of the data (e.g. proportion responding by category)?

In planning a measurement exercise, the aim should be to clearly specify the desired output, and the data items required to produce this, before considering question design. This will involve, at a minimum, defining the measures to be used and the breakdowns and cross-classifications required. In many cases, particularly if multivariate analysis is proposed, more detailed information may be required.

Questionnaire design

Once a clear set of outputs has been identified, based on the analysis required to meet users' needs, it will be possible to make specific decisions about survey design, including the most appropriate survey vehicle, collection period, units of measurement and questionnaire design. These decisions should flow logically from the process of working down from users' needs through analysis and output. The remainder of this chapter sets out a strategy for making these decisions. This includes both specific proposals for how a national statistical agency might approach the measurement of trust and more general information that can be used in a wider range of circumstances.

What other information should be collected: Covariates and analytical variables

All potential uses of trust data require some understanding of how trust varies with respect to other variables. This applies whether the goal is understanding the drivers of trust – which requires understanding the causes of change – or monitoring trust over time and across countries – which requires factoring in changes in demographics, in order to understand whether a given change in the average trust of a community is due to changes in levels of trust reported by different demographic groups or in the shares of these groups

in society. It is therefore imperative to consider not only how best to measure trust *per se*, but also what other measures should be collected alongside measures of trust for analytical purposes.

A need for additional information to aid in interpreting and analysing results is not unique to trust. Most statistical measures are collected alongside, at the least, basic demographic data. Demography matters to trust measures just as much as it does to labour-market statistics. There are pronounced differences in average levels of trust across a range of different demographic groups. Better educated and higher income groups typically have a higher level of generalised trust, while women generally have slightly lower levels (Alesina and La Ferrara, 2001; Soroka, Helliwell and Johnston, 2003; Helliwell and Wang, 2010). Generalised trust increases with age, but at a declining rate, while being a member of an ethnic group with a history of discrimination is associated with lower levels of generalised trust (Alesina and La Ferrara, 2001).

The precise range of covariates to collect alongside measures of trust will depend on whether the focus is on interpersonal trust or institutional trust, and on the research question being examined. Despite this, it is possible to provide some general guidelines on the most important information that should be collected alongside measures of trust. Most of the covariates described below are regularly collected by national statistical agencies, and international standards for their collection do exist. No attempt is made here to detail how these variables should be collected, and it is assumed that existing standards apply.

Demographics

Demographic variables cover the basic concepts used to describe the population being measured and to allow the analysis of how outcomes vary by population group. Including a range of demographic measures in any attempt to measure trust is of utmost importance, in particular for the following breakdowns:

- ***Age*** of the respondent: This should be provided in single years, if possible: age bands, while allowing for some cross-classification, are less useful both because they allow less flexibility with respect to the groups examined, and because they do not facilitate analysis of age as a continuous variable.
- ***Gender*** of the respondent.
- ***Marital status:*** This could be either the legal marital status of the respondent, including whether the respondent is widowed, divorced or separated, or the *de facto* status, including whether the respondent is living as married even if not legally married.
- ***Household type:*** This refers to a classification of the respondent's household unit, including whether the respondent is single or living with a partner, and whether children are present.
- ***Children:*** The number and age of children in the respondent's household, along with the relationship to the respondent.
- ***Household size:*** The number of people living in the respondent's household. Household size is a distinct concept from family size, as more than one family unit can live in a dwelling. It is important particularly with respect to income, in order to calculate the equivalised income available to household members for consumption purposes.[2]
- ***Geographic information:*** While privacy concerns may prevent the release of detailed geographical information relating to the respondent, estimates should be disaggregated

by some broad level geographic regions such as urban and rural, capital city, states/provinces, etc. Geo-coding allows for merging with other datasets also containing geo-codes, such as environmental data or other characteristics of the areas where people live. This is particularly important for the analysis of institutional trust, as variation between different local authorities provides a way to link governmental performance with trust in institutions without requiring comparable cross-country data.

- ***Migration status/Country of birth/Year of arrival:*** Migration status, such as permanent residence, citizenship, etc., and/or country of birth of the respondent. This information is especially important for generalised trust, both because migration status may have implications for generalised trust within a country and because the country of origin affects trust on an ongoing basis (Algan and Cahuc, 2010). Where space permits, the country of birth of parents is also of high interest, as it allows the analysis of trust for second-generation migrants.
- ***Ethnic identification:*** The ethnic identity or identities of the respondent may be of high policy importance in diverse societies. Ethnic identification is known to affect levels of interpersonal trust (Alesina and La Ferrara, 2001) and is also of interest from the perspective of institutional trust (Statistics New Zealand, 2015).

In addition to the demographic measures identified above, which can be considered essential, a number of additional demographic variables may also be useful. The precise relevance of these may, however, vary depending on national circumstances and the research priorities being considered.

- ***Language:*** Beyond the primary language of the respondent, it may be desirable, in some cases, to collect information on other languages spoken at home. Proficiency in the main language of the country where the survey is taking place may also be important for some purposes.
- ***Urbanisation:*** The classification of the area where the respondent lives in terms of the degree of urbanisation.[3]

Social and economic outcomes

In addition to basic demographic information, which allows for the identification of *who* trusts within society, there is also a need for a range of variables relating to wider social and economic outcomes that can be used to address the questions *what drives trust?* and *how does trust affect other valued outcomes?* While the variables listed below do not represent an exhaustive list of potential covariates, they identify those that are of the greatest interest and which are likely to be the most intensively used.

- ***Income:*** Income is of high interest since trust (both interpersonal and institutional) is known to vary with income (Alesina and La Ferrara, 2001). While both personal and household income are of interest, household income is of the highest priority, as it drives consumption possibilities and living standards, and therefore likely plays a major role in shaping trust. Ideally both pre- and post-tax income should be collected, as this would allow for the analysis of the impact of the tax and transfer system in shaping the trust of individuals.
- ***Savings and wealth:*** The relationship between savings behaviour, wealth accumulation and trust is of high interest, as trust in financial institutions, in the regulation of the financial sector and in other individuals can be expected to play a major part in peoples' savings decisions. Savings, in turn, is an important policy issue from the perspective of

both macroeconomic decisions and retirement income policy. Because savings and wealth data impose a significant respondent burden if collected in detail, the specific choice of variables used will depend on the context. In a general household survey, priority should be given to relatively simple questions such as the forms of wealth held (e.g. bank account, shares, private pension, land) or wealth brackets rather than detailed estimates of total wealth.

- ***Employment status:*** Employment status is a standard control variable for the analysis of social and economic outcomes. Although often omitted from the analysis of trust outcomes – because employment adds little explanatory power to other socio-economic control variables (e.g. Algan and Cahuc, 2013; Soroka, Helliwell and Johnston, 2003; Helliwell and Wang, 2010) – employment status is nonetheless important in order to allow assessing the independent effect of trust on other outcomes that are also affected by employment (e.g. income). The measurement of employment status is covered by existing (ILO) guidelines related to labour-market statistics, which provide the relevant standards for questions in surveys related to trust.[4]
- ***Educational attainment:*** Educational attainment is a major driver of trust at the individual level (Helliwell and Wang, 2010; Algan and Cahuc, 2013), and it is an essential control variable to include when trust data are collected. Existing standards covering the collection of data on educational attainment provide a clear basis for measurement in this area.
- ***Health status:*** While there is relatively little research on the effect of the respondent's health conditions on trust, information on health status is of interest from the perspective of investigating the impact of trust on health (e.g. Ginn and Arber, 2004; Stafford et al., 2005). While much of this analysis has been conducted looking across countries, including health status measures in surveys containing trust data would allow analysing the link between health and trust at the individual level. Although it is difficult to measure health status accurately in a household survey, there are a number of widely used survey instruments available, ranging from the health status descriptions included in the World Health Organisation's survey (WHO, 2012) through to more specialised question modules such as the GHQ-22 for mental health (Goldberg et al., 1978). More recently, a joint activity between the UNECE, WHO and Eurostat (the Budapest Initiative) has led to a *Survey Module for Measuring Health State* (UNECE, 2013), which provides a standard set of survey questions for measuring health status.
- ***Social contact and networks:*** Social contact and the social networks that exist between people have an obvious connection to trust. The existence of this link is well supported by the literature, which identifies many links between social behaviour and levels of trust at both the individual and cross-country level (Helliwell and Wang, 2010). Although there are currently no international standards on the collection of measures of social contact, measures relating to how much time people spend in social interaction, the frequency of social interaction, the number of people that respondents interact with and the nature of their relationships are all of high interest for analysis of the drivers of trust.
- ***Civic engagement and governance:*** Following Putnam (1993) and Fukuyama (1995), the links between trust, civic engagement and governance are of high policy interest. However, although institutional trust is often regarded as one of the key measures of civic engagement and governance in its own right, the relationship between trust and other measures of civic engagement and governance remains unclear. Levels of

volunteering, including the time spent volunteering and the form of volunteering, are all relevant, as is participation in other forms of civic activity such as voting. Beyond this, information on the perceived quality of governance and on satisfaction with the basic services provided by public institutions is of particular relevance when collected alongside measures of institutional trust. Including information on the local government institutions that the respondent is associated with is also of high importance, in that it allows variation in local government performance to be used to examine the relationship between the quality of governance and trust.

- ***Personal security and victimisation:*** Experiences of victimisation have an intuitive link to levels of interpersonal trust, and trust in the judicial system is correlated with levels of crime (see Chapter 2) at the country level. At the individual level, the relationship is less clear, partly due to lack of data. Measures of both experienced victimisation and perceived safety should be collected, as is already done in victimisation surveys, given that the two approaches to measuring personal security often produce different results.
- ***Subjective well-being:*** Subjective well-being and interpersonal trust are strongly correlated, both at the individual level (Helliwell and Wang, 2010) and across countries (Boarini et al., 2013). Both Boarini et al. (2013) and Helliwell et al. (2014) find evidence of a strong relationship between aspects of institutional trust and measures of subjective well-being. Incorporating measures of subjective well-being into surveys containing trust measures allows for analysis of the impact of trust on subjective well-being and is important in attempts to measure well-being more broadly.[5]
- ***Religion*** is commonly used as a control variable in studies looking at trust (e.g. Soroka et al., 2003; Algan and Cahuc, 2013; Nannestad, 2008), both because of the potential impact of religious differences on levels of generalised trust in society and because of the potential role of religious participation in establishing social cohesion within groups.

4.3. Survey and sample design

One important distinction between measures of trust and many of the measures typically included in official statistics is that trust measures will almost invariably need to be collected through sample surveys. In contrast to many economic or population statistics, no administrative database could produce information of this sort without, in effect, incorporating survey questions in the administrative process.[6] Thus, survey and sample design are fundamental to producing valid and reliable measures of trust.

It is not the role of this chapter to provide detailed guidelines on sample frames and sample design. These are specialist areas in their own right, and excellent guides exist for data producers who are seeking advice on these technical aspects of data collection (United Nations Statistical Division, 1986). However, in survey design, as in other aspects of design, form should follow function. When trust is the specific goal of measurement, this has implications for survey design. This section discusses some of the most significant considerations for the measurement of trust with respect to the target population, when and how frequently data should be collected, what collection mode should be used, and what the most appropriate survey vehicle is.

Target population

The target population of a survey describes the complete set of units to be studied. A sample survey will generally attempt to achieve a representative sample of the target

population. However, in some circumstances the target population may be focused on one or more specific sub-groups of the total population from which the sample is drawn. It may also specify sub-populations that the survey should include in a broader sample representative of the total population. For example, the total population might be *all persons aged 15 and over living in private dwellings in a specified area*. However, the target population might also specify men and women as sub-populations of interest, requiring the sampling frame to accommodate distinct analysis of these two groups. More generally, sub-groups are often defined by characteristics such as age, gender, ethnicity, employment status or migrant status.

Some surveys using households as the measurement unit rely on a single respondent (such as the head of household) to provide responses for the household as a whole. This approach cannot be used for measures of trust, since the cognitive process of evaluating and responding to questions on whether a person or institution is trusted is very different to that of providing an estimate of another householder's educational attainment or labour-market status. Responses to questions on trust are intrinsically personal, and consequently the unit of measure must be the individual. This implies that the sampling frame must produce a representative sample of individuals as if all individuals are personally interviewed. While this will typically not be an issue for surveys where the individual is the primary unit of analysis, some household surveys may require an additional set of individual weights to derive individual estimates. Surveys where the response is on the basis of *any responsible adult*, or which allow *proxy respondents* when the selected person is not present at the time of the interview, will be problematic in this regard.

The target age group for measures of trust will also vary with respect to the goals of the research programme. For example, in the context of research on retirement income policies, it may be appropriate to limit the target population to persons aged 65 or older. In general, however, measures of trust would usually be collected for all the adult population (aged 15 and older). There is little research, and currently relatively little policy demand, for data on trust by children. For this reason, issues relating to children's levels of trust are not covered here.

People not living in private households

One population group that may be of high policy interest, but which is not typically covered in household surveys, is people not living in private households. This group includes people living in institutions, including prisons, hospitals or residential care facilities, as well as people with no fixed residence, such as the homeless. These groups raise two issues with respect to the measurement of trust. The first problem is common to all attempts to collect statistical information on such groups – i.e. that such people tend to be excluded from standard sample frames used for household surveys. This means that, at a minimum, specific data collection efforts will be required based on a sample frame designed to cover the relevant institutions. In some cases, such as the homeless, it may be difficult to develop any statistically representative sampling approach at all.

A more significant challenge facing the measurement of trust is that many of the people in the relevant groups may not be able to respond. This is particularly the case for people institutionalised for health-related reasons that affect mental functioning (including people with some mental illnesses or with physical illnesses limiting the ability to communicate, and the very old). In these cases, it is not possible to collect information on trust from the respondent. Proxy responses, which might be appropriate for some types of data (income, marital status, age), are not valid in the case of measuring trust.

Frequency and duration of enumeration

The frequency with which data are collected, the enumeration period, typically involves a trade-off between survey goals and available resources. All other things being equal, more frequent collection of data will improve the timeliness of estimates available to analysts and policy makers, and will make it easier to discern trends in the data over time. More frequent enumeration, however, is more costly, both in terms of the resources involved in conducting the data collection and in terms of the burden placed upon respondents. It is therefore important that decisions about the frequency of data collection are made with a clear view to the relationship between the timeliness and frequency of the data produced and the goals of the data collection.

It is not possible to provide specific guidelines for how frequently measures of trust should be collected that cover every contingency, since the range of possible data uses is large and the frequency at which data are needed will vary depending on the intended use and on the type of measure in question. However, some general advice can be provided. Aggregate measures of trust generally change only slowly over time (Nannestad, 2008). This reflects the relatively slow movements in most of the drivers that affect trust and the fact that trust – particularly interpersonal trust – is associated with long-standing cultural values (Putnam, 1993; Uslaner, 2002, 2008).

Small changes in measures of trust might appear to suggest that such measures do not need to be collected frequently. However, small absolute changes also mean that standard errors tend to be large relative to observed changes. A number of observations are therefore needed to distinguish between a trend over time and noise in the data. Box 4.2 illustrates this point. For this reason, despite (or indeed, because of) the relatively slow rate of change in trust data, it is desirable that measures are collected on a regular and timely basis. In particular, for the purposes of monitoring well-being and for measuring trends in levels of social capital, an annual time series should be regarded as the essential minimum in terms of frequency of enumeration. (It should also be noted that frequent or rolling sample-surveys increase the possibilities for identifying the causal impacts of other factors, because it will be possible to analyse how changes in trust follow changes in other factors.)

Box 4.2. **Identifying appropriate trust measurement frequency**

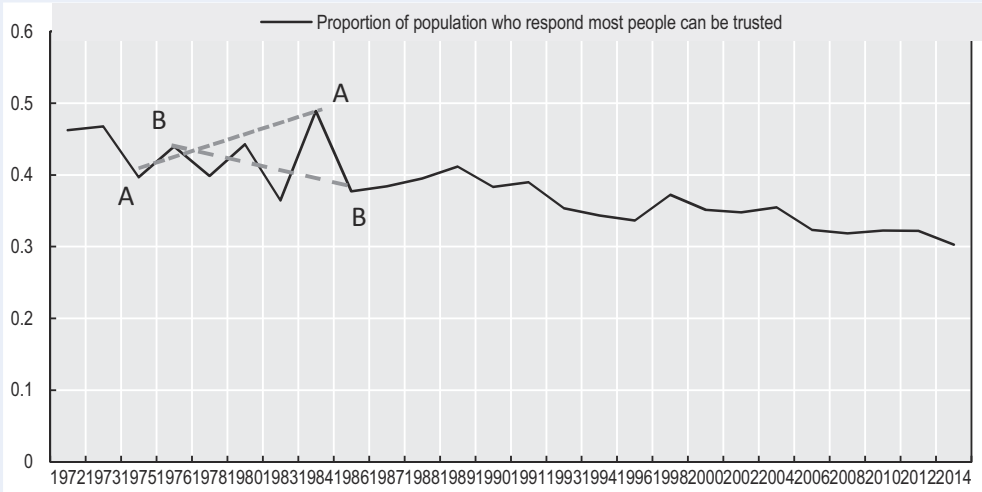
If measures of trust change only slowly over time (Nannestad, 2008), then it might seem logical to conclude that data need only be collected infrequently. After all, why go to the trouble and expense of collecting many observations that are close together and show little change? This, however, ignores the impact of measurement error associated with collecting data from sample surveys. Figure 4.2 shows changes over time in levels of generalised trust in the United States from the American General Social Survey (GSS). Over the period covered by the figure, data collection took place at least every two years, and for many periods data were collected annually. The initial sample size was approximately 1 000, rising to about 2 000 during the 1980s and to over 4 000 from 1994.

Across the whole time period covered by the GSS, generalised trust has trended downwards, from about 45% of the sample to just over 30%. There is little evidence of any periods during this time frame when the trend has been significantly different. However, there are periods – particularly during the 1970s and early 1980s – when the GSS sample size was smaller, implying that the volatility of the series around the trend was relatively high.

Box 4.2. **Identifying appropriate trust measurement frequency** *(cont.)*

Over this period, despite the fact that the trend in levels of trust is for a steady but relatively gentle decline, the picture of trust that emerges is very sensitive to the available data.

Figure 4.2. **Generalised trust in the United States, 1972-2014**



Source: NORC (2014), General Social Survey (database), *https://gssdataexplorer.norc.org/*.

StatLink *http://dx.doi.org/10.1787/888933583842*

If only two data points were available – 1975 and 1984 – then the GSS would have shown a substantial *increase* in trust over time (line A_A). On the other hand, if the two data points used for comparison were 1975 and 1986, then the GSS would have shown a rapid *decrease* in trust (line B_B). It is only because the GSS data were collected on an annual or biennial basis that it is possible to observe that the actual development over the course of the 1970s is neither a smooth increase nor a sharp fall, but instead a gradual decline associated with higher volatility in the measure. The significance of these measurement issues is smaller in the second half of the period covered by the GSS, where the data are not as volatile. This reflects the fact that, with larger sample sizes, the frequency of measurement is less important to identifying trends over time. However, even allowing for this, there is still a strong case for using additional resources to collect data more frequently rather than focusing on infrequent larger samples: frequent surveys increase the timeliness of data, reducing the gap between a policy issue being identified and the availability of new data.

Duration of enumeration

The duration of the enumeration period is likely to be important for measures of trust. Unlike measures of educational attainment or marital status, for which it does not usually matter at what point during the year the data are collected, the precise timing of the collection period might have an impact on measured trust. Although there is little hard evidence on the magnitude of these timing effects, comparisons with other self-reported measures (e.g. Deaton, 2011) suggest that this can be substantial.

The fact of being sensitive to the point in time at which the data are collected is not unique to measures of trust. Many core labour-market statistics, for example, have a pronounced seasonality, and published statistics usually adjust for this. However, in order to

produce the information required for seasonal adjustments, data need to be acquired over the course of a whole year. Ideally, enumeration of trust data would take place over a full year, and would include all days of the week, including holidays. This would ensure that measures of trust provide an accurate picture across the whole year. Where a year-long enumeration period is not possible, enumeration should, as far as is possible, be spread proportionately over all the days of the week.

Holidays (and to some degree, periods of annual leave) are problematic in that they tend to be distributed unevenly over the course of the year, and they affect how people feel. This has been found to have an impact on how respondents report some other subjective measures such as life evaluation (Deaton, 2011); similar effects might also affect measures of trust. Thus, if enumeration cannot be spread over a whole year, there is a risk that an incidence of holidays during the enumeration period that is greater or lesser than normal might bias the survey results. For this reason, in surveys collected with relatively short enumeration periods it is essential to check the impact of including data collected during any holidays. While it may not be necessary to omit data collected during holidays from output if the impact is negligible or weak, the available evidence on the magnitude of some holidays suggests that it is important to test for potential bias from this source. What constitutes a holiday will also need to be considered with respect to the context in which the survey is conducted.[7]

Sample size

Large samples are highly desirable in any survey, as they reduce the standard error of estimates and allow both a more precise estimate of trust as well as a greater degree of freedom with respect to producing cross-tabulations and an analysis of results for population sub-groups. With measures of trust, sample size is particularly important because of the relatively small changes in trust associated with many areas of analytical interest. While a very significant shock such as the global financial crisis of 2008 can cause large shifts in some trust measures, this is not always the case, and many changes are quite small. For example, although the Eurobarometer measure of the share of the population in Greece, Ireland and Spain expressing trust in government fell, on average, by 20% between 2008 and 2013, the proportion of the population in the United Kingdom expressing trust in government fell by only 3.5% during the same period.

Although it is impossible to give precise guidelines for what is an appropriate sample size, some general criteria can be noted. Most of the factors that should be taken into account in the planning of any survey also apply when collecting information on trust. Available resources, respondent burden, sample design (e.g. a stratified sample will have a different sample size compared to a random sample with the same objectives, all other things being equal), the anticipated response rate and the required output will all influence the desirable sample size. The need for sub-national estimates, in particular, will play an important role in determining the minimum required sample. However, it is also worth noting that, again, all other things being equal, measuring trust is likely to require a somewhat smaller sample than measuring concepts that affect only a small part of the population, such as unemployment or victimisation.

Survey mode

Surveys can be carried out in a number of different modes. Because the mode of collection influences survey costs and respondent burden, as well as inducing mode effects in responses, the choice of survey mode is an important decision when collecting data. The

two modes most commonly used to collect information on self-reported measures are Computer-Assisted Telephone Interviewing (CATI), conducted by an interview over the telephone, and Computer-Assisted Personal Interviewing (CAPI), where the interviewer is personally present when recording the data. Computer-Assisted Self-Interview (CASI) surveys can occur in the presence of an interviewer; when the interviewer is on hand but the respondent enters their own data into a computer questionnaire; or without an interviewer present, as in the case of an internet survey. For some purposes, traditional self-completed surveys are still likely to be relevant.

As outlined in Chapter 3, there is good evidence that the collection mode has a significant impact on responses to trust questions. In general, CAPI has the advantage that the interviewer can build rapport with the respondent. CASI with the interviewer present offers similar advantages. Conversely, CATI interviews do not allow for the same degree of interviewer interaction with the respondent, and the rapport between interviewer and respondent may be lower. CASI has the additional advantage that it may address issues associated with social desirability bias in questions on interpersonal trust and, especially, institutional trust.

As with other features of survey design, the choice of the survey mode is influenced by a variety of factors, including resource constraints. However, the balance of evidence suggests that, where resources permit, CASI with the interviewer present (i.e. a CASI/CAPI mix) is likely to produce the highest data quality. This is due in part to the rapport that interviewers can build in face-to-face situations. However, CAPI also provides the opportunity to use show cards, which CATI lacks. Show cards that include verbal labels for the scale end-points are particularly valuable when collecting information on trust, where the meaning of the scale end-points changes between questions, as this can impose a significant cognitive burden on respondents (ONS, 2012).

In terms of data quality, CASI/CAPI with show cards should be considered best practice for collecting trust data. The presence of an interviewer allows for a strong rapport to be built with the respondent, while show cards help with data quality. The confidentiality provided by CASI, through the self-administration of the survey, should help address respondent reluctance to provide accurate answers to sensitive questions, especially if additional confidentiality assurances are provided (see Chapter 3). Where other modes are used, it is important that data producers collect information to enable estimating the impact of mode effects. National statistical agencies, in particular, should consider experimentally testing the impact of the survey mode on responses to the core measures of trust and publishing the results along with those from CATI or CASI surveys.[8]

Survey vehicles

As discussed earlier in this chapter, analytical interest in measures of trust is commonly focused on the interaction between trust and other social and economic outcomes. Also, in most cases, trust measures are relatively simple and easy to collect. Even a relatively comprehensive approach to measuring trust is likely to be on the scale of a module that could be added to existing surveys rather than requiring a whole survey questionnaire in itself. A key question to consider then is which survey vehicles are most appropriate to the task of measuring trust.

Where trust, governance or social capital is the key focus of interest, it may be appropriate to build a special topic module around this theme. This is especially the case

where the use of trust data focuses on measuring social capital or on evaluating governance. Beyond this, however, trust measures are relevant to a number of different types of survey.

It is currently impossible to provide definitive guidance on this issue, because the range of household surveys conducted – even among national statistical agencies – varies significantly from country to country. Box 4.3 outlines what a system of statistics on well-being might look like and the potential role of trust measures within this framework. The system of well-being statistics proposed by Box 4.3 builds on the recommendations of the Commission on the Measurement of Economic Performance and Social Progress (Sen, Stiglitz and Fitoussi, 2009) to provide a data infrastructure for measuring the different elements of well-being and their main drivers. However, it is possible to identify the roles that different survey vehicles can play in collecting trust data within existing constraints.

General social surveys

Not all national statistical agencies run general social surveys and, among those that do, the content and focus vary considerably. Some agencies, such as the Australian Bureau of Statistics, focus their general social survey primarily on measures of social capital and social inclusion, while others rotate modules on different topics between waves (Statistics Canada) or are explicitly multidimensional (Statistics New Zealand). All three approaches, however, make such surveys ideal for the inclusion of measures of trust. All three of the general social surveys cited above (Australia, Canada and New Zealand) currently include measures of interpersonal and institutional trust. Surveys with rotating content, such as the Canadian General Social Survey, offer the opportunity for a trust module that can collect information in some depth if this is determined to be a priority. Surveys with a wider focus, such as the New Zealand General Social Survey, are particularly valuable in that they allow for the analysis of the joint distribution of trust and of a wide variety of other topics, including material conditions and other aspects of quality of life (joint distribution refer to measures that are "joined up" at the individual unit-record level and hence allow to see how multiple outcomes of interest are spread across individuals). Regardless of whether a specific trust module is conducted as part of a general social survey, it would be very desirable that at least the core module of trust measures be collected in all general social surveys.

Victimisation surveys

Victimisation surveys collect information on the level and distribution of criminal victimisation in a society. They are intended to answer questions such as how much crime takes place, what are its characteristics, whether the level of crime is changing over time, who are the victims or the people at greater risk of becoming one, and how do perceptions of safety relate to the actual risk of victimisation (UNECE, 2010). The interaction between victimisation, perceptions of safety and trust is of high interest, both from the perspective of understanding how victimisation affects trust in others and in order to better understand the relationship between trust in the justice system (police, courts, etc.) and victimisation. The inclusion of a module of trust questions in victimisation surveys should be a high priority.

Special topic surveys

Many national statistical agencies run one-off or periodic surveys on special topics that are intended to explore an issue of interest in greater detail than would be possible

through a question module in a regular survey. Because the content of such a survey can be tailored to the topic in question, such surveys are excellent vehicles for exploring aspects of trust in more depth. In particular, special topic surveys may be useful for examining institutional trust and its drivers in more detail than is possible through a survey with broader coverage, or for examining the relationship between different measures of trust (e.g. *Trustlab*, see Chapter 2). However, because of the "one-off" nature of such surveys (or the long periodicity associated with such surveys when they are repeated), special topic surveys are less appropriate for monitoring trust over time.

Panel surveys

Panel surveys follow the same people over time, re-interviewing them in each wave of the survey. Because of this, panel surveys are able to examine questions of causality in a way that is not possible with cross-sectional surveys. Both the German Socio-Economic Panel (GSOEP) and Britain's Understanding Society (formerly the Household Panel Survey) have included questions on trust for some time. The GSOEP has been the basis for some important experiments on the validity of trust measures (Naef and Schupp, 2009).

Other surveys

In addition to the surveys listed above, there are a range of other surveys where trust measures may be of some relevance to specific research questions, but for which it is harder to make a case for including trust in the core questionnaire. Health surveys are a core part of most national statistical systems, collecting information on health outcomes, their determinants and health behaviours. Although not a primary source for aggregate data on interpersonal or on institutional trust more generally, there is a clear case for collecting information on trust in the health-care system within these surveys. Beyond this, information on interpersonal trust and on some of the wider aspects of institutional trust may be of relevance to a health survey as potential determinants of health outcomes or as drivers of health behaviours.

A similar rationale may be made for including measures of trust in surveys that relate to education. Questions on trust have been included in both the international PISA and PIAAC surveys, reflecting the potentially important role that education has in shaping social attitudes such as trust (Borgonovi and Burns, 2015) and the fact that aspects of institutional trust may be a significant driver of educational outcomes.

Trust questions also have potential relevance in time-use surveys. These surveys typically involve respondents filling out a diary, collecting detailed information on the respondent's allocation of time to different activities and with whom those activities were undertaken. Because of the wealth of detailed information that they contain on the nature and extent of interpersonal contact, time-use surveys are ideal instruments for investigating the relationship between interpersonal trust and different forms of social contact, if the relevant trust questions are included. Currently most time-use surveys do not include questions on trust.

Wealth and savings surveys represent another vehicle where trust questions are potentially important. In particular, trust in financial institutions has obvious relevance for understanding savings behaviour. Beliefs about the trustworthiness of financial institutions are a major determinant of savings. Interpersonal trust is also of potential relevance as a determinant of savings behaviour.

4.4. Questionnaire design

Designing a suitable questionnaire is an iterative process involving questionnaire designers, those responsible for determining survey content, and data users. A questionnaire designer must balance the cognitive burden on the respondent, a limited time budget for the survey, and the need to have a questionnaire that is clear and comprehensible and that flows well, with different (and often competing) data needs. It is neither possible nor desirable for this chapter to provide a single questionnaire on trust that users should implement. Instead, the intent of this section is to provide a set of tools to support the development of surveys containing questions on trust, rather than to prescribe a single approach to its measurement.

Box 4.3. **Integrating trust into a system of well-being statistics**

A statistical framework brings together a conceptual framework relating to the concept of interest, the measurement instruments required to quantify it, and the statistical infrastructure needed to ensure that data are collected in a way consistent with statistical quality standards. The best-known statistical framework is the System of National Accounts (SNA), but population statistics and labour force statistics are equally supported by coherent statistical frameworks that underpin the production of high-quality, timely and internationally comparable data. Conversely, no commonly accepted statistical framework exists for household well-being statistics. While several national statistical offices have made steps in this direction, or signalled their intent to do so in the future (Bycroft, 2011; Dupré and Di Meglio, 2014), these efforts are still in their infancy.

From a measurement perspective, the primary characteristic of well-being statistics is their multidimensionality. Well-being statistics need to cover subjects as diverse as income, employment, health status, social contact, environmental quality, governance and subjective well-being. This breadth of scope places a significant burden on national statistical offices. Beyond this, however, are several additional challenges. Well-being statistics need to provide information on the distribution of outcomes, both in continuous terms and for specific population sub-groups. This implies a relatively large sample size. In addition, many areas of well-being, such as health status or social contact, do not reduce easily to a single measure: hence there is a need for dedicated in-depth measurement for some areas, using a range of different instruments. Finally, well-being measures need to be "joined up" at the individual unit-record level, so as to allow analysis of the joint distribution of outcomes (Sen, Stiglitz and Fitoussi, 1999) and of their drivers.

These issues raise conflicting priorities for well-being statistics. On the one hand, the need for joined-up statistics across multiple outcomes could be met by a single household survey covering all topics briefly. On the other hand, the need for in-depth measures on specific topics points towards drawing data from more detailed surveys focused on a single area, such as health status or the labour market. Both these priorities need to be balanced against the limited financial resources available to national statistical offices and the high demand for data on other topics.

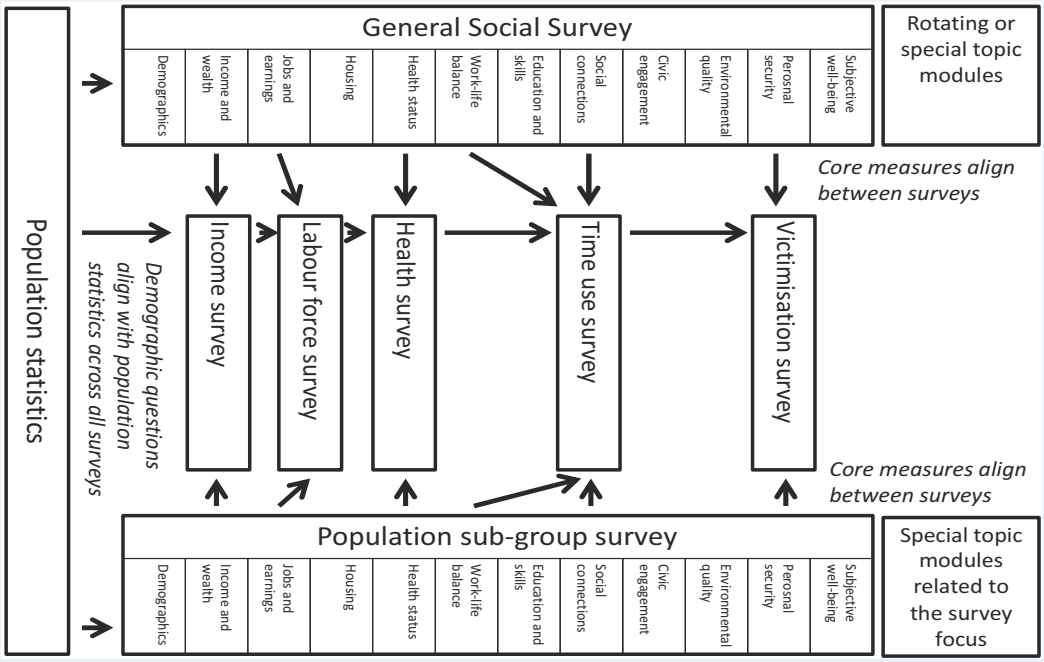
One way to reconcile these conflicting demands is to make the greatest possible use of existing data sources by integrating them to provide a coherent portfolio of well-being statistics that can be joined up via a set of core indicators for each outcome domain. Standardisation of the core social variables would allow for cross-cutting datasets to be linked to more detailed data sources that provide more specialised information on a particular topic. Such a system is outlined in Figure 4.3.

In Figure 4.3, a General Social Survey collects information on each domain of well-being along with core demographic information and some analytical variables. To lower respondent burden and cost, a GSS cannot collect in-depth information on each well-being domain, so data collection focuses on two or three core indicators for each domain. Beyond collecting information on a limited set of core indicators for each domain, one central role of a GSS is to allow cross-classifying respondents based on alternative outcomes, so as to allow links to surveys on specific subject matters that allow more in-depth analysis. These subject matter surveys

Box 4.3. **Integrating trust into a system of well-being statistics** *(cont.)*

mostly already exist in national statistical systems (e.g. labour force surveys, household income surveys, health interview surveys, time-use surveys, victimisation surveys) and can be linked to a GSS by a set of common core variables collected in both surveys.

Figure 4.3. **A system of well-being statistics: Conceptual overview**



Source: Fleischer, L., C. Smith and C. Viac (2016), "A Review of General Social Surveys", *OECD Statistics Working Papers*, No. 2016/09, OECD Publishing, Paris, *http://dx.doi.org/10.1787/bb54d16f-en*.

StatLink http://dx.doi.org/10.1787/888933583861

Such an approach also allows for cross-cutting surveys focused on specific population groups of interest such as older people or specific ethnic groups. Specific surveys focused on sub-populations may be needed either because the policy issues relevant to the group require additional information that is not relevant to society more widely, or because the sub-population in question is sufficiently small that it is not well-reflected in a population survey. Using a common set of core indicators allows both comparing the position of the population sub-group of interest to that of the population as a whole and making linkages with more detailed subject matter surveys.

In the context of a system of social statistics, trust plays a key role among the core social indicators. In particular, institutional trust measures represent one of the strongest candidates for a core measure of civic engagement that can be collected in a general social survey. Similarly, generalised trust is the strongest candidate for a core measure of social capital and is highly relevant to measuring social connections more widely. The core measures included in the question module recommended by these Guidelines (Annex 2) include the key questions that would ideally form the basis for trust-related core indicators in a system of well-being statistics.

Some general guidance on issues affecting the inclusion of measures of trust in a survey is provided below. In particular, the issues of question placement and translation are discussed in detail. This is accompanied by a set of prototype question modules that

questionnaire designers should adapt to the specific conditions in which they are working. This section also describes the rationale behind the question modules and provides an explanation of the template used to describe them. The question modules are attached to these Guidelines as Annex 2 (A to E).

Question placement

Question order and the context in which a question is asked can have a significant impact on responses to subjective questions such as those on trust (see Chapter 3). Although measures of trust are not uniquely susceptible to such effects – question order and context will impact upon all survey responses to some extent – the effect is likely to be relatively large in the case of trust.

In general, question order effects occur not because the question appears early or late in the questionnaire *per se*, but because of the contextual impact of the immediately preceding questions. Thus, the key issue is to identify the most effective way to isolate questions on trust from the contextual impact of preceding questions. The most direct way of managing contextual effects of this sort is to put trust questions as early in the survey as possible. Ideally, such questions should come immediately after the screening questions and household demographics that establish respondent eligibility to participate in the survey. This practice almost eliminates the impact of contextual effects and ensures that those that cannot be eliminated in this way are consistent from survey to survey.

However, there are two reasons why this cannot be a general response to the issue of dealing with contextual effects. First, there will be instances when questions on trust are added to well-established surveys. In these conditions, changing the flow of the questionnaire would impose significant costs in terms of both resources and data quality. Introducing questions on trust early in such a survey might ensure that contextual effects do not impact the trust questions, but this would come at the expense of creating significant contextual effects for the following questions. Second, in cases where there are several types of subjective question in the survey (such as questions on trust and on subjective well-being), they obviously cannot all be first.

With these factors in mind, four key recommendations can be provided with regard to the placement of trust questions in surveys:

- *Place important questions on trust in the core section of the survey*. Although it is not possible to place trust questions at the start of every survey,[9] the effect of bias due to context effects on analysis can still be limited if trust questions are included in a fixed portion of the survey questionnaire. While this does not eliminate bias, it will not affect analysis of differences in levels across population groups or over time.
- *Avoid placing the trust questions immediately after questions that are likely to prime respondents with regard to trust or that respondents might use as a heuristic for determining their response to the trust question*. This would include questions on social contact, victimisation or political beliefs or any questions suggesting risk or insecurity. The best questions to precede subjective questions are relatively neutral factual demographic questions.
- *Make use of transition questions to refocus respondent attention*. One technique that has been used with respect to subjective well-being, and which may also be useful with questions on trust, is to use a transition question designed to focus the respondent's attention away from issues that might unduly influence their response. Deaton (2011), in an analysis of subjective well-being data, reports that the introduction of such a question in the Gallup

Healthways Well-being survey in 2009 eliminated over 80% of the impact of a preceding question on politics on the subsequent life evaluation measure.[10] A similar approach might also be applied with respect to trust questions. However, it is important to consider the risk that transition questions might introduce their own context effects. For example, drawing attention to a respondent's personal life may lead them to focus on personal relationships or family rather than on strangers when answering subsequent questions about interpersonal trust. Development of effective transition questions should be a priority for future work.

- *Use of introductory text to distinguish between question topics*. Well-worded text that precedes each question or topic can serve as a buffer between measures of trust and sensitive questions. However, there is little hard evidence on the degree of effectiveness or optimal phrasing of such introductory text. A standard introductory text has been included in each of the prototype question modules included as Annex 2 to this chapter. This text is based on what is believed to be best practice. Consistent use of it should help reduce context effects (and will eliminate bias caused by inconsistent introductory text). Further cognitive testing or experimental analysis of the impact of different types of introductory text would, however, be of high value.

Question order within and between trust questions

Questions on trust can potentially be affected just as easily by previous trust questions as by questions on other topics. This has implications for the structure of question modules on trust (particularly where more than one aspect of trust is addressed) as well as for the presentation of questions within modules and for whether it is advisable to include several questions that address very similar topics (see Chapter 3).

In terms of ordering question modules themselves, overall the evidence suggests that moving from the general to the specific is the best approach. This has different implications for interpersonal and institutional trust. For interpersonal trust, it suggests that a question on generalised trust should be placed ahead of more specific questions relating to who is trusted. It also suggests that questions on interpersonal trust should be asked ahead of questions on institutional trust, which tend to be more specific and are also likely to have a significant priming effect, because they direct the respondent's attention to political issues that are known to have a powerful priming effect in other contexts (Deaton, 2011).

For institutional trust, the issue of question order is less clear-cut, in that there is no overarching concept of institutional trust that could be collected before the more detailed questions relating to trust in specific institutions. However, it is possible that priming effects may follow from the questions on trust in different institutions. This has two main implications. First, institutional trust questions should proceed from the easier to the more difficult for respondents to answer; for example, questions on trust in parliament or in the judicial system should precede questions on institutions such as the United Nations or the European Commission. Second, where comparisons between levels of trust in different institutions are important, the order in which questions are presented should be randomised for each respondent; when this is not possible for the whole sample, pilot testing should involve randomisation of question order so that the size of any bias in measured trust in different institutions is known.

Finally, asking two questions about a very similar construct can be confusing for respondents, leading them to provide different answers because they think that different answers are required of them. This means that including several very similar questions

about interpersonal trust, for example, could elicit respondents to answer these questions differently than when each question is presented in isolation. This implies that it is important to have consistency both in the number of measures used to assess a given construct and in the order in which those measures are asked in the survey.

Translation

The exact question wording used when collecting information on trust can matter a lot for responses. As discussed in Chapter 3, a standardised approach to question wording is important for comparisons over time or between groups. This is relatively straightforward where all surveys are in the same language. However, international comparisons or studies in multilingual countries raise the issue of translation. This is a non-trivial matter in the case of trust. For example, a number of English language surveys distinguish between *trust* in a person and *confidence* in an institution (e.g. Gallup World Poll, World Values Survey). However, this distinction does not exist in other languages. In French, for example, both *trust* and *confidence* would be translated as *confiance*, suggesting that the implicit conceptual distinction in the GWP survey questionnaires does not translate well in other languages.

Potential issues arising from translation cannot be eliminated entirely, but they can be managed through an effective translation process. An example of good practice in the translation of survey questionnaires is provided by the *Guidelines for the Development and Criteria for the Adoption of Health Survey Instruments* (Eurostat, 2005). Although focused on health survey instruments, the framework for translation presented in that report has broader applicability, and is highly relevant to the measurement of trust. The Eurostat health survey guidelines identify four main steps in the translation procedure:

- *initial or forward translation* of the questionnaire from the source document to the target language
- *independent review* of the translated survey instrument
- *adjudication* of the translated survey instrument by a committee to produce a final version of the translated survey instrument
- *back translation* of the final version of the translated survey instrument into the source language.

Most of the best-practice recommendations identified by Eurostat for health surveys also apply with respect to the measurement of trust. It is desirable that the initial translation be carried out by at least two independent translators who have the destination language as their mother tongue and who are fluent in the source language. Translators should be informed about the goal of the study and be familiar with the background, origin and technical details of the source questionnaire as well as with the nature of the target population. The reviewer at stage 2 should be independent from the translators but will ideally have a very similar set of skills. Both the reviewer and the translators should be on the adjudication panel, along with an adjudicator whose main area of expertise is the study content and objective. As with any survey design, cognitive interviewing and field testing should be undertaken, with the results reviewed before the full survey goes into the field.

Back translation is somewhat controversial in the literature on survey translation, with some experts recommending it and others not (Eurostat, 2005). The effect of back translation is generally to shift the focus onto literal translation issues rather than the conceptual equivalent of the original instrument. In the case of the measurement of trust, back translation is strongly advised, due to the sensitivity to question wording of trust measures (see Chapter 3).

Choice of questions

The choice of which questions to use is of critical importance for measuring trust. Different questions capture different types of trust and, as discussed in Chapter 3, the precise question wording can have a non-trivial impact on results. In selecting questions to incorporate into existing survey vehicles, statistical agencies face trade-offs between the time taken to ask any new questions, the potential impact of new questions on responses to existing questions, and the added information gained from the new questions. These trade-offs will come under severe scrutiny when the survey in question refers to an important and well-established concept (e.g. household income or unemployment).

When selecting trust questions themselves, there is also a trade-off to manage between using existing questions from the literature that will enable reasonable comparability with previous work, and modifying questions or response formats already used in existing surveys in light of what has been learned about good practice – including the evidence described in Chapter 3. The approach adopted in this chapter is to recommend tried-and-tested questions from the existing research first and foremost. Where a variety of approaches have been used in the past, the rationale for selecting between these is explained. Finally, where there is a case for making small alterations to the question wording based on the evidence in Chapter 3, modifications are proposed.

For statistical agencies already using trust measures in their surveys, a crucial question will be whether the potential benefit of using improved measures, and/or more internationally comparable ones, outweighs the cost of disrupting an established time series. This is a choice for individual statistical agencies to make, and it will depend on a number of factors including what the current and future intended use of the dataset is, how drastic the change may be, and how long the time series has been established for. It is recommended that any changes to existing questions are phased-in using parallel samples, so that the impact of the change can be fully documented and examined. This will enable insights into the systematic impact of changes in methodology and provide agencies with a potential method for adjusting previous datasets (e.g. Deaton, 2011).

In recognition of the different users' needs and resources available to statistics producers, this chapter does not present a single approach to gathering information on trust. Instead, five question modules are attached to the Guidelines as Annex 2 (A to E). Each question module focuses on a distinct approach to measuring trust. Question Module A, the "core module", contains the core measures for which international comparability is the highest priority. These are the measures for which the evidence for their validity and relevance is greatest, the results are best understood, and the policy uses are the most developed. Of the five question modules included in Annex 2, Module A is unique in that national statistical agencies are encouraged to implement it in its entirety. When this is not possible, the single primary measure outlined in the module should be used at a minimum. Modules B to E are focused on different approaches to measuring trust. These modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires and would like to know what has already been tested in household surveys.

The five modules are listed in the following; those which it is highly recommended that national statistical offices implement as they are are highlighted as *recommended,* in order to distinguish them from the modules intended as a *resource* for data producers of all types who are developing more detailed questionnaires.

Recommended

- A. Core Measures

Resource

- B. Evaluation
- C. Expectations
- D. Experiences
- E. Experiments

A. Core Measures

The core measures module contains those questions for which there is the strongest evidence of validity and for which the policy relevance is the strongest, and which therefore are most apt to achieve some degree of international harmonisation. Within the core measures, a distinction is made between the ***primary measure***, which is the single question that is intended to form the baseline for international comparisons and which should be regarded as the highest priority for inclusion in any attempt to measure trust, and the ***supplementary questions***. Unlike all of the other question modules included in these Guidelines, the core module is intended to be used without significant amendment. Beyond this, it is also recommended that national statistical offices try to implement the core module (or at least the primary measure) in at least one regular household survey. If a general social survey is conducted on a regular basis by a national statistical office, then this forms the ideal vehicle for the inclusion of the core measures. However, the core measures are also intended to be used to connect different surveys. For example, in the situation where a country has both a general social survey module that includes a wide range of trust measures and a survey of migrants where trust is of interest but is not the main focus, both surveys should include the primary measure at a minimum. This would allow analysis of the migrant survey using the primary trust measure to be informed by the more detailed picture of trust available from the general social survey.

The core measures of trust proposed in Annex 2 consist of five questions organised into three groups. The first question (A1) is the ***primary measure*** of generalised trust in others and is intended to be the baseline measure used for international comparisons. This is supplemented by an additional question on limited interpersonal trust (A2) and by three questions on institutional trust (A3, A4 and A5). Taken together the core module questions provide a minimal coverage of the main dimensions of trust.

The primary measure of trust is the essential minimum that should be included in all surveys containing trust measures. It is focused on generalised trust, both because the evidence for the validity of generalised trust measures is greater than is the case for measures of institutional trust (see Chapter 2) and because generalised trust has wide policy relevance in that it is the best available candidate for a measure of the stock of social capital driving a wide range of outcomes. Generalised trust hence is of crucial importance for the measurement of a society's well-being, and it is also closely related to levels of institutional trust (Rothstein and Uslaner, 2005).

In selecting the primary measure of trust, important issues needed to be addressed. First, should the measurement approach be based on evaluations or on expectations? On the evaluative side, the Rosenberg question on generalised trust[11] and its variants have been used widely and are well-understood. The evidence for the validity and reliability of the

Rosenberg question is strong, and the question has been applied in a wide range of different countries and cultural contexts (see Chapter 2). The main alternative to the Rosenberg question is the "lost wallet" question[12] used in the Gallup World Poll and some other contexts (Helliwell and Wang, 2010). This is an expectations question that focuses on how the respondent expects other people to behave. Although with a much narrower history of use, it has been used in a number of different countries and there is relatively good evidence on its validity. Its main interest compared to the Rosenberg question is that the anecdotal nature of the question may be intuitively appealing to respondents.

However, when compared to the Rosenberg question, the lost wallet question has three main drawbacks. First, although the evidence for its validity is relatively good, it is still much less thoroughly understood than the Rosenberg question and does not have nearly such a long history of use. Second, there is little evidence linking answers to the lost wallet question and laboratory experiments. In contrast, the Rosenberg question has been the subject of a large experimental literature, and the relationship between it and experimental behaviour is much better understood. Finally, the Rosenberg question is much more adaptable to different forms of scale labelling. In particular, it is relatively easy to adapt it to a numerical scale, while the lost wallet question loses much of its intuitive appeal if asked in conjunction with a numerical scale. For this reason, an evaluative question based on the Rosenberg question is recommended by these Guidelines as the primary measure of generalised trust.

The precise choice of scale and wording to be used is the second major issue associated with the choice of primary measure. In its original form, the Rosenberg question is dichotomous, and the scale labels relate to slightly different concepts. One response (*people can generally be trusted*) focuses unambiguously on trust, while the other response (*you can't be too careful*) combines notions of trust and caution (Naef and Schupp, 2009). Following Chapter 3, a 0 to 10 end-labelled scale is preferred to a dichotomous scale, mainly because the greater range of response options allows the question to capture more variation in responses between individuals. A scale from 0 to 10 is thought to represent the upper end of the viable range of response options in terms of the ability of respondents to make meaningful use of the full range of scale values. In addition, based on the evidence reviewed in Chapter 3 on the impact of including a reference to caution on response distributions, the primary question recommended here focuses exclusively on trust, dropping the implied reference to caution in the original Rosenberg question (*you can't be too careful*).

The primary measure recommended by these Guidelines is similar to questions on generalised trust included in a number of official surveys (e.g. the Australian General Social Survey, the Canadian General Social Survey's limited trust questions, the New Zealand General Social Survey) in terms of focusing exclusively on trust, rather than on caution. It is most similar to the question used in the New Zealand General Social Survey, which also uses a 0 to 10 end-labelled scale (see Annex 1 for questions from these surveys).

The second question in the core module is a measure of limited trust. This measure is included as a complement to the primary measure rather than an alternative. The measure is taken from the World Values Survey module on the extent of trust, but has been modified to adopt a similar format to the other questions in module A. While limited trust is of less policy interest than is the case for generalised trust or institutional trust, information on limited trust can provide a useful context for interpreting the analysis of other trust measures.

Three questions on institutional trust form the final part of the core measures. As discussed in Chapter 2, the evidence for the validity and reliability of measures of

institutional trust is less strong than is the case for measures of interpersonal trust. For this reason, these measures should be considered to have a more experimental nature than the other core measures, particularly when compared to the primary measure. However, these questions do represent best practice for collecting information on institutional trust, and variants of them have been used in official surveys in a number of countries. As with the questions on interpersonal trust, the questions use a 0 to 10 end-labelled numerical scale. The scale labels are conceptual absolutes (*not at all/completely*).

The core module questions cover three institutions. This is based on the outcome of factor analysis of a much wider range of institutions reported in Chapter 2 (Box 2.1), which indicates that responses to questions on institutional trust reflect three underlying dimensions: politics, law and order, and non-governmental institutions. In the case of the first two questions (A3 and A4), the specific institutions selected are those that have the highest factor loading for each of the first two dimensions (politics, law and order), reflecting the strongest individual correlation with the underlying dimension. These questions relate to the country's parliament and the police. The third question on institutional trust (A5) does not refer to the third dimension identified in the factor analysis (non-governmental institutions), but instead provides additional information on the country's public institutions that are not explicitly political in the appointment of their staff and mandate, focusing on the civil service. (The civil service loads, albeit less strongly, in the same factor as political institutions.)

B. Evaluation

Evaluative questions collect information on how a respondent judges a situation. They have a cognitive and reflective component and represent the respondent's view of how things are now. They do not seek information on past experiences, nor are they forward-looking in the manner of expectations questions (see sub-section C below). The questions included in the evaluation module of Annex 2 fall into three groups, each of which expands on the measures in the core module by providing more detail on an aspect of trust. They are not intended to be all used together, but rather form a resource for developing survey questions focused in more detail on specific aspects of trust. Questions B1 to B4 have been extensively validated in their current format, and should be used unaltered, while the other questions have a greater degree of flexibility.

The first block of four questions (B1 to B4) is taken from Naef and Schupp (2009) and provides a measure of generalised trust that is believed to be more precise and reliable than the traditional Rosenberg question. In particular, the Naef and Schupp module is designed to distinguish between trust and caution and also asks separately about *strangers* and *people* more generally. Questions B1 and B3 focus on trust, while questions B2 and B4 focus on caution. Similarly, questions B1 and B2 ask about people generally, while questions B3 and B4 ask about strangers specifically. Thus, the scope of the Naef and Schupp module is similar to the original Rosenberg question but allows for the different aspects of the original question to be examined separately. Because the Naef and Schupp module consists of four questions, the measure also has greater discriminatory power than the dichotomous scale originally proposed by Rosenberg. The Naef and Schupp scale has the additional advantage that it has been tested in the German Socio-Economic Panel survey and has been validated directly against experimental behaviour through the trust experiment (see Chapter 2). It is for this reason that the original four-point scale is left unaltered here rather than adopting a 0 to 10 scale as used elsewhere.

Questions B5 to B8 focus on who is trusted (i.e. limited trust). They are intended to complement the primary measure of trust by providing more detail on trust in different groups in society. This will be of particular interest where different social groups may exhibit systematic differences as to what types of people they have in mind when responding to the primary measure. The questions themselves are based loosely on a question module from the World Values Survey (Delhey, Newton and Welzel, 2011) but have been adapted to bring them in line with the methodological guidelines set out in Chapter 3. Although Delhey, Newton and Welzel propose an adjusted measure of generalised trust based on this set of questions, it is not clear that their composite measure outperforms the core measure proposed here in capturing the most significant outcomes thought to be associated with generalised trust (Algan and Cahuc, 2013). For this reason, the questions are mainly recommended for use in situations where the details of trust in specific groups are of high importance.

The final block of questions in the evaluative module (B9 to B18) focuses on institutional trust. This set of questions is an extension of the three questions included in the core module to cover a wider range of institutions; they are drawn from the World Values Survey and from wave 6 of the European Social Survey. Based on evidence from analysis of the OECD Trust Database and other similar analyses (Uslaner, 2002; Naef and Schupp, 2009; Schneider, 2016), it appears that respondents often fail to fully distinguish between similar institutions. As previously noted, much of the variation in responses to institutional trust questions derives from three underlying dimensions relating, respectively, to politics, law and order, and non-governmental institutions. Nonetheless, this set of questions is intended to provide a framework for investigating attitudes to specific institutions. The three questions on institutional trust from the core module (A3, A4 and A5) are repeated here as questions B9 to B18 and are designed to stand on their own. If they are implemented alongside the full core module, then questions B9, B13 and B16 can be omitted.

C. Expectations

The main alternative to obtaining people's evaluations with respect to trust is to ask them about their expectations of the behaviour of other people and institutions. This is the approach taken by the questions included in this module. Unlike evaluations, expectations questions are forward-looking and hypothetical. In one sense, this represents a weakness, as respondents may be less able to provide accurate answers about hypothetical situations than they are about their current beliefs and evaluations. However, this is partly offset by the fact that the anecdotal nature of expectations questions may make them more concrete and easier for respondents to answer.

Questions C1 and C2 relate to interpersonal trust. These are drawn from a series of questions used in the Gallup World Poll and have also been applied in other contexts, including the Canadian General Social Survey. Question C1 focuses on trust in neighbours (limited trust), while C2 refers to trust in strangers (generalised trust). Although there is no evidence that expectations questions perform better than evaluations, there is evidence that expectations questions of this sort have a reasonable level of validity and have broadly similar correlates (Helliwell and Wang, 2010). These questions have been included as a potential complement to the primary measure, to be considered only in circumstances where there may be value in using an additional methodological approach to measurement.

The bulk of the expectations questions (C3 to C10) relate to aspects of institutional trust. These questions are not direct substitutes for the questions contained in the core module or in the evaluation module, because they do not focus on specific institutions but instead on

the main factors that are believed to influence how respondents form their views on institutional trust. Specifically, the questions are intended to collect information on each of the five dimensions of the OECD Trust Framework developed in the context of the OECD Trust Strategy to assess ways of restoring trust in institutions (Box 4.3). The questions are organised into groups of three, with each group relating to behaviours associated with a specific set of drivers of institutional trust from the OECD Trust Framework: question C3 relates to responsiveness; C4 to reliability; C5 to openness; C6 and 7 to fairness; and C8 to C10 to integrity of public institutions.

Box 4.4. **The OECD Trust Framework**

The public management literature has acknowledged the importance of trust as a key source of the legitimacy of public institutions and its importance to lowering transactions costs in interactions between people, businesses and institutions (Fukuyama, 1995). However, the evidence about what drives trust in institutions is scattered. Some analyses have described trust in institutions as depending on the congruence between people's (and businesses') preferences (i.e. their interpretation of what is right and fair and what is not) and the perceived functioning of government (Bouckaert and Van de Walle, 2003). Other authors have drawn a distinction between "trust in competence", i.e. whether the functioning of institutions matches people's expectations about the competencies of those steering them, and "trust in intentions", which captures whether institutions act in a way that is perceived by people as ethical and fair (Nooteboom, 2007). These distinctions are furthered by Bouckaert (2012), who distinguishes between the "logic of consequences", where trust is derived causally from outcomes, and the "logic of appropriateness", where trust is based on the values of integrity and transparency.

Based on the distinction between the outcomes of an action and the intentions that guided it, the OECD Public Governance Committee endorsed a framework for assessing people's trust in public institutions. The framework deconstructs trust in institutions into two key components: *Competence* and *Values*. Within each component, relevant dimensions that are amenable to policy change are identified based on the common threads in the literature (McKnight, Choudhury and Kacmar, 2002) and on the OECD update of this evidence (OECD, 2017). The five dimensions identified are: *Responsiveness, Reliability, Integrity, Openness* and *Fairness*. Table 4.1 describes the framework and the aspects considered within each policy dimension.

Table 4.1. **The OECD Trust Framework**

Trust component	Government mandate	Concern affecting trust	Policy dimension
Competence: Ability of governments to deliver citizens the services they need at the quality level they expect	Provide public services	- Access to public services, regardless of social/economic condition - Quality and timeliness of public services - Respect in public service provision, including response to citizen's feedback	Responsiveness
	Anticipate change, protect citizens	- Anticipation and adequate assessment of evolving citizens' needs and challenges - Consistent and predictable behaviour - Effective management of social, economic and political uncertainty	Reliability
Values: The drivers and principles that inform and guide government action	Use power and public resources ethically	- High standards of behavior - Commitment against corruption - Accountability	Integrity
	Inform, consult and listen to citizens	- Ability to know and understand what government is up to - Engagement opportunities that lead to tangible results	Openness
	Improve socio-economic conditions for all	- Pursuit of socio-economic progress for society at large - Consistent treatment of citizens and businesses (vs. fear of capture)	Fairness

Source: OECD (2017), *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264268920-en*.

StatLink http://dx.doi.org/10.1787/888933584222

Box 4.4. **The OECD Trust Framework** *(cont.)*

Finally, attempts have been made to operationalise the framework by developing questions similar to those widely used to measure interpersonal trust. The questions used to operationalise the OECD framework refer to a stereotypical situation and inquire about its expected outcome. The following is an example of a situational question for the integrity dimension: *If a high-level politician was offered a well-paid job in the private sector in exchange for political influence, do you think that he/she will reject the job?*

The potential advantage of using expectations-based questions to collect information on the drivers of trust in institutions is that such information may be less affected by attitudes to the government of the day, and are less likely to be dominated by information on the three main dimensions of trust in institutions (politics, the justice system and non-governmental organisations) outlined in Chapter 2 (Box 2.1). Despite this, questions C3 to C17 should be regarded as highly experimental. While they have been trialled by the OECD as part of *Trustlab*, and are based on a review of the available literature (OECD, 2017), they have only a limited record of use. In including these questions in these Guidelines, the intent has been to provide a framework that other producers of trust data could use for developing more detailed questions. The hope is that these questions will encourage the development of better measures of institutional trust that can fill the current gap with respect to evidence on the convergent validity of institutional trust measures.

D. Experiences

Although not used as widely as evaluative or expectations measures, several questions on trust exist that relate to the respondent's experience of trusting or trustworthy behaviour. In contrast to many other areas of statistics, where asking about people's actual experiences is the norm, it is difficult to identify unambiguous examples of specifically trusting behaviour, which makes it difficult to develop reliable questions. Nonetheless, a small number of questions of this sort have been used, either to validate other types of trust measure or simply because the behaviour in question was of interest in its own right. The experience module (module D) in Annex 2 contains five questions: the first three of these relate to interpersonal trust, while the last two relate to institutional trust.

Questions D1, D2 and D3 are taken from Naef and Schupp (2009) and are intended to capture the respondent's prior history of trusting behaviour. They focus on examples of behaviour where the respondent has demonstrated a willingness to accept a risk based on a positive expectation of others' behaviour. D1 and D2 are focused on limited trust, while D3 is more relevant to generalised trust. The Naef and Schupp questions used here are, in turn, derived from questions developed by Glaeser et al. (2000), but with a number of minor changes. The most significant change is to remove wording from the Glaeser questions that related specifically to the university students who were the target of the original questionnaire.[13] In both Naef and Schupp and in Glaeser et al., the three questions were used together to build an index of past trusting behaviour. Using the questions in this way is one of the possible applications of the question module, but such questions can also be used individually as measures of trust. While questions D1 to D3 have been used before, they have not received the same degree of scrutiny or testing that has been applied to the main evaluative or expectations-based questions included in these Guidelines. This is because the main use for experience questions has been to validate other approaches to measuring trust.

However, this does not preclude more sophisticated experiential questions on trust from being developed; providing a starting point for developments of this sort is the second main reason for including these questions in Module D.

In the case of institutional trust, there are no questions designed specifically for the purpose of collecting information on respondents' experiences of trusting behaviour with respect to institutions. However, this remains an interesting and potentially fruitful area for development. Two questions are included in Annex 2 (D4 and D5). Both are taken from the Gallup World Poll and relate to examples of behaviour that demonstrate trust in the integrity of public institutions (voicing an opinion to an official and signing a petition). Compared to experiential questions about interpersonal trust, this represents a very limited range of experiences, and it clearly represents a large gap in trust measurement. Additional question development in this area has the potential to bring real benefit.

One point worth further exploration relates to the potential for measuring trust in institutions by analysing the gap between people's reported experiences and institutions' records of actual behaviour. An example of this would be to look at the gap between instances of victimisation reported to police and the victimisation rate measured through victimisation surveys. A high level of trust in the police would presumably be correlated with a high rate of victimisation instances actually reported to them, while a lack of trust would be correlated with a lower rate. Questions of this type are not included in these Guidelines, since measuring trust in this way would require analysing existing data for other outcomes.

E. Experiments

The final question module included in Annex 2 of these Guidelines differs from the previous four in that it focuses on experimental protocols for measuring trust rather than on traditional survey questions. The reasons for including an experimental module in the Guidelines are twofold. First, experimental measures of trust have been implemented in the context of large-scale household surveys: for example, Naef and Schupp (2009) describe how a Trust Game has been implemented in the German Socio-Economic Panel survey, while the OECD's *Trustlab* aims to develop a cross-country measurement tool based on nationally representative samples including both traditional survey questions and a range of experimental games. The ability of experimental games to be incorporated into household surveys means that, even if this measurement approach is still unusual for national statistical offices, it falls within the scope of the Guidelines and may be of relevance to some data producers – particularly those outside the official statistical system. The second reason for including experimental protocols in these Guidelines is the importance of generating experimental data that could be used for validating survey-based measures of trust. Even if experimental measures of trust are not widely used, it is potentially helpful to users of these Guidelines to include details of the main experimental games that have been used to validate the main survey questions.

There are a wide range of experimental protocols designed to shed light on different aspects of human behaviour. Only two are included in the experimental module proposed here: these are, respectively, the Trust Game and the Dictator Game. The Trust Game is the standard laboratory approach to measuring trust and trustworthiness, and provides the main experimental data related to trust. The Dictator Game is included here as a complement to the Trust Game. A number of studies have suggested that information from

the first mover in the Trust Game conflates both altruism and trust (Cox, 2004). The Dictator Game, by way of contrast, collects a relatively pure measure of altruism (Kahneman, Knetsch and Thaler, 1986), which can then be used to control for the impact of altruism in the Trust Game. The version of the Trust Game set out here is more relevant to generalised trust than limited trust, as it is anonymous.

As discussed in Chapter 2, the Trust Game was first developed by Berg, Dickhaut and McCabe (1995) and has been used extensively since then. The protocol for the Trust Game included in module E of Annex 2 is a version of the Berg, Dickhaut and McCabe Trust Game adapted by the OECD for implementation in *Trustlab*, in the context of nationally representative surveys. This approach involved considerable time and energy for designing a version of the Trust Game that could be implemented without substantive changes (beyond translation) in a range of different countries and cultural contexts. Similarly, the implementation of the Trust Game described in module E assumes that the game participants are responding to a CASI questionnaire rather than in a laboratory setting. This is particularly important because in a laboratory game participants can be matched with other players in real time, while in a CASI questionnaire players in the game are likely to be matched post-hoc. This requires a wider range of responses to be captured, particularly for the second-moving player.

The Dictator Game is easier to implement in a CASI setting, since only one player makes a substantive choice. As with the Trust Game, the version of the Dictator Game implemented here is that developed by the OECD for *Trustlab* (see Box 2.4). This has the advantage that the game has already been reviewed and tested in a number of different countries and is suitable for inclusion in a large-scale survey.

Beyond the two games included in module E, there are a range of other experimental approaches to measuring trust or elements of human behaviour relevant to trust that could be trialled. These include other experimental games, such as the Public Good Game, which measures people's willingness to contribute to a public good in a situation where public and private incentives differ; experimental games concerned with measuring risk attitudes; and the use of experimental techniques other than games, such as implicit association testing. Implicit association testing is a technique used by psychologists to collect information in cases where social desirability bias is likely to make it difficult to get an honest answer from respondents (e.g. on racism). The OECD and Sciences Po have trialled the use of implicit association tests to measure aspects of institutional trust in *Trustlab*, and there is considerable interest in these techniques among experimental economists, political scientists and sociologists (see Intawan and Nicholson, 2017). While it is currently too early to include a developed question module in these Guidelines, further developments in these areas are likely to have a significant impact on approaches to measuring trust over the coming decade.

Question Templates

The five question modules are attached to these Guidelines as Annex 2. Each question module is presented in the same format, containing a common set of headings that outline the objectives of the module (i.e. what kind of information it is trying to gather), its content, the origin of the questions, how the data from the module should be presented, background information for interviewers, and the detailed question wording.

Objectives

The objectives provide an overview of the intended role of the module and how it is supposed to be used. They provide a succinct description of the purpose of the module and of the rationale behind its scope and contents.

Description

A description of the contents of each question module outlines the role of each of the questions in the module with respect to the module's objectives. The description is intended to assist users to identify which questions they wish to use in the event that they choose to implement only part of the module.

Origin

Questions included in each module are drawn from existing sources and remain unchanged wherever possible to maximise comparability with previous work. However, some items have been modified to a greater or lesser extent where several forms of the same question have been used in the literature and/or where there are clear grounds for making small changes in item wording or response scales, for example based on the evidence in Chapter 3. This section indicates the source of the questions and notes whether they have been altered.

Completion Time

This gives an estimate of the time required to run the entire module.

Output

The output section contains basic information on the production of standard tables and measures from the question block. This information is not exhaustive, but is intended to provide some basic guidelines for data producers. Such guidelines are important not only to assist producers in presenting the data appropriately but also to provide context for why the questions are framed in the way that they are.

A number of the question blocks are intended to produce multi-item measures of trust derived from the survey questions. This section provides details on the construction of these multi-item measures and how they should be reported.

Guidelines for interviewers

The quality of any survey data is heavily influenced by the attitude of the respondents to the questions that they are being asked. Although the evidence is overwhelming that measures of trust are not regarded by respondents as particularly sensitive or difficult to answer (particularly when compared to questions on some other commonly asked topics, such as income), better-quality information is likely to result if interviewers understand what information is being collected and how it will be used, and they are able to communicate this clearly to respondents. This enables interviewers to answer queries from respondents on why the information is important or on what concept the question is trying to elicit from them.

The guidelines for interviewers contained in this module are not intended as a substitute for the more extensive notes and/or training that would normally be provided to interviewers in the process of preparing to conduct a household survey. However, they do provide a basis from which users of the module can develop their own more substantive guidelines.

4.5. Survey implementation

How a survey is implemented is crucial to its effectiveness. A carelessly implemented survey will result in low-quality and unreliable data regardless of the quality of the underlying questionnaire. In general, the features relevant to the effective implementation of any household survey also hold for those collecting information on trust. These Guidelines make no attempt to provide a detailed discussion of best practice in survey implementation, for which high-quality standards and guidelines already exist (United Nations Statistical Division, 1984). However, there are several points where the specific nature of measures of trust raises survey implementation issues that are worth noting.

Interviewer training

Interviewer training is crucial to the quality of responses in any survey. However, the measurement of trust raises additional issues, because the subject matter may be unfamiliar to interviewers. This is, ironically, particularly so for national statistical agencies with a permanent force of field interviewers. Although a body of trained interviewers will generally contribute to higher response rates and better responses, interviewers may struggle with questions if they cannot explain adequately to respondents why collecting such information is important and how it will be used. Academic literature, anecdotal evidence and feedback from national statistical office staff suggests that some trust questions – particularly those related to trust in political institutions – may be of particular sensitivity, and that respondents may in some situations be concerned about the uses of such data (Schneider, 2016). In some cases, respondents may fail to understand why a public agency might want to collect this type of information.

To manage risks around respondent attitudes to questions on trust, it is imperative that interviewers are well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used. This is essential in order for interviewers to build a rapport with respondents, which can be expected to improve compliance by respondents and the quality of responses. While the interviewer guidelines contained in the question modules provide some crucial information specific to each set of questions, a more comprehensive approach should draw on information on the validity and use of measures of trust (Chapter 2) and on the analysis of trust data (Chapter 5).

Ethical issues

Evidence suggests that measures of trust are relatively non-problematic for respondents to answer. Item non-response rates are generally low for measures of both interpersonal and institutional trust (see Chapter 2). In general, item-specific non-response rates for interpersonal measures are similar to those for marital status, education and labour market status, and much lower than for measures of income. This suggests that, in general, such questions are not perceived as problematic by respondents. Non-response rates for measures of institutional trust are a little higher than is the case for interpersonal trust, but still remain low in absolute terms.

Best practice suggests that statistical providers should consider how to manage the risks associated with questions that are distressing to respondents. Although it is important not to overstate the risks – there is little in the way of evidence suggesting that trust questions pose a significant risk in this respect – such issues should be dealt with effectively. A complicating factor is that it might not be evident at the time of the interview

whether a respondent has been affected by the questioning. One approach to managing this factor, used by the ONS (2012), is to distribute a leaflet at the time of the interview giving respondents information on the purpose of the survey and reiterating the confidentiality of the data collected. The leaflet would also contain information for distressed respondents about where to seek help.

Coding and data processing

The coding of information on trust is generally straightforward. In general, numerical scales should be coded as numbers, even if the scale bounds have labels. Much analysis of trust data is likely to be quantitative and will involve manipulating the data as if they were cardinal. Even for fully-labelled response scales (such as the *yes/no* responses that apply to many questions), it is good practice to code the data numerically as well as in a labelled format in order to facilitate use of the microdata to produce summary measures. Responses of *don't know* and *refused to answer* should be coded separately from each other, as the differences between them are of methodological interest.

Normal data-cleaning procedures include looking for obvious errors such as the transposition of numbers by data coders, duplicate records, loss of records, incomplete responses, out-of-range responses or failure to follow correct skip patterns. Some issues are of particular relevance to trust data. In particular, where a module comprising several questions with the same scale is used, data cleaning should also involve checking for response sets, which occur when a respondent provides identical ratings to a series of different items. For example, a respondent may answer "0" to all ten domain evaluative questions from module E. This would typically suggest that the respondent is not, in fact, responding meaningfully to the question and is simply moving through the questionnaire as rapidly as possible. Such responses should be treated as a non-response and discarded. In addition, interviewer comments provide an opportunity to identify whether the respondent was responding correctly, and a robust survey process will make provision for allowing such responses to be flagged without wiping the data record.

Finally, it is important to emphasise that much of the value from collecting measures of trust comes from microdata analysis. In particular, analysis of the joint distribution of trust and other outcomes cannot usually be accomplished through the secondary use of tables of aggregate data. Because of this, a clear and comprehensive data dictionary should be regarded as an essential output in any project focusing on trust. This data dictionary should have information on the survey methodology, sampling frame and correct application of survey weights, as well as a description of each variable (covering the variable name, the question used to collect it and how the data are coded). If a variable is collected from only part of the survey sample due to question routing, this should also be clearly noted in the data dictionary.

4.6. Conclusion

Key points made in this chapter are as follows:

Planning for the measurement of trust

- Decisions about what to measure should always be grounded in a clear understanding of user needs. Important questions to consider include: i) What are the policy questions?; ii) Is the trust content being proposed appropriate to respond to the policy questions?; iii) Does the measure proposed allow monitoring changes over time or comparing

population groups?; iv) What population groups are of greatest interest to the user?; v) Does the user's interest lie in comparing outcomes of different groups or in understanding the relationship between different aspects of trust?; vi) Is the user's primary interest in generalised trust, limited trust or institutional trust? If the focus is on the latter, which institutions are of primary interest?; vii) What are the frequency requirements of the users to monitor changes over time?; and viii) What within-country comparisons are required, such as geographic level?

- It is imperative to consider not only how best to measure trust *per se*, but also what other measures should be collected alongside measures of trust for analytical purposes. These should include: i) Age; ii) Gender; iii) Marital status; iv) Household type; v) Presence of children; vi) Household size; vii) Geographic information; and viii) Migration status/ Country of birth/Year of arrival.
- In addition to the demographic measures identified above, which can be considered essential, a number of additional variables may also be useful: i) Language; ii) Living in urban/rural areas; iii) Income; iv) Wealth; v) Employment status; vi) Educational attainment; vii) Health status; viii) Social contact and networks; ix) Civic engagement and governance; x) Personal security and victimisation; xi) Subjective well-being; xii) Ethnic identification; and xiii) Religion.

Survey and sample design

Sampling

- Responses to questions on trust are inherently personal, and consequently the unit of measure must be the individual. This implies that the sampling frame must produce a representative sample of individuals or households as if all individuals are personally interviewed.
- In general, measures of trust would be collected for the entire adult population (aged 15 and older).

Frequency of data collection and duration of enumeration

- It is not possible to provide specific guidelines for how frequently measures of trust should be collected that cover every contingency, since the range of possible data uses is large and the frequency at which data are needed will vary depending on the intended use and on the type of measure in question.
- For the purposes of monitoring well-being and for assessing trends in social capital, an annual time series should be regarded as the minimum in terms of frequency of enumeration.

Duration of enumeration

- The duration of the enumeration period (i.e. the period of time over which information is collected) is important for measures of trust. Unlike measures of educational attainment or marital status, for which it does not usually matter at what point during the year the data are collected, the precise timing of the collection period might have an impact on measured trust.
- Ideally, enumeration of trust data would take place over a full year and would include all days of the week, including holidays. This would ensure that measures of trust provide an accurate picture for the whole year. Where a year-long enumeration period is not

possible, enumeration should, as far as is possible, be spread proportionately over all days of the week.

Sample size

- Large samples are highly desirable in any survey: they reduce the standard error of estimates and allow both a more precise estimate of trust as well as a greater degree of freedom when producing cross-tabulations and analysis of results for population sub-groups. With measures of trust, sample size is particularly important because of the relatively small changes in trust associated with many areas of analytical interest.

Survey mode

- In terms of data quality, CASI/CAPI with show cards should be considered best practice for collecting trust data. The presence of an interviewer allows for a strong rapport to be built with the respondent, while show cards help with data quality.
- The confidentiality provided by CASI sections to the interviewing should help address respondent reluctance to provide accurate answers to potentially sensitive questions.
- Where other modes are used, it is important that data producers collect information to enable the impact of mode effects to be estimated. National statistical agencies, in particular, should consider experimentally testing the impact of the survey mode on responses to the core measures of trust and publish the results along with any results from CATI or CASI surveys.

Survey vehicle

- Where trust, governance, or social capital is the key area of interest, it may be appropriate to build a special module focused specifically on trust. This is especially the case where the use of trust data focuses on measuring social capital or on evaluating governance.
- As trust measures are of analytical interest in a broad range of different contexts, a limited range of trust questions can usefully be included in a wide range of different types of survey.

Question placement

- Important trust questions should be included in the core section of the survey. Although it is not possible to place trust questions at the start of every survey, the effect of bias due to context effects can be limited if trust questions are included in a fixed portion of the survey questionnaire. While this does not eliminate bias, it will not affect analysis of differences in levels across population groups or over time.
- Trust questions should not be placed immediately after questions that are likely to prime respondents with regard to trust, or that respondents might use as a heuristic for determining their response to the trust question. This includes questions on social contact, victimisation, political beliefs, risk or insecurity. The best questions to precede trust questions are demographic questions.
- Transition questions should be used to refocus respondent attention. However, it is important to consider the risk that transition questions might introduce their own context effects. For example, drawing attention to a respondent's personal life may lead them to focus on personal relationships rather than on strangers when answering subsequent questions about interpersonal trust. Development of effective transition questions is a priority for future work.

- Introductory text should be used to distinguish between question topics. Well-worded text preceding each question or topic can serve as a buffer between measures of trust and sensitive questions. Further cognitive testing or experimental analysis of the impact of different types of introductory text would be of high value.

Question order

- In terms of ordering question modules, evidence suggests that moving from the general to the specific is the best approach. For interpersonal trust, a question on generalised trust should be placed ahead of more specific questions relating to limited trust. Also, questions on interpersonal trust should be asked before questions on institutional trust, which tend to be more specific and are also likely to have a significant priming effect.
- Questions on trust in institutions should proceed from better-known institutions to more obscure ones. Where comparisons between levels of trust in different institutions are important, the order in which questions are presented should be randomised for each respondent. If this is not possible for the whole sample, pilot testing should involve the randomisation of question order so that the size of any bias in measured trust is known.

Translation

- Initial translation should ideally be carried out by at least two independent translators who have the destination language as their mother tongue and who are fluent in the source language. Translators should be informed about the goal of the study and be familiar with the background, origin and technical details of the source questionnaire, as well as with the nature of the target population. As with any survey design, cognitive interviewing and field testing should be undertaken, with the results reviewed before the full survey goes into the field.

Questionnaire design

- These Guidelines provide five prototype question modules for the measurement of trust. Module A contains a set of ***core measures***, which include a single primary measure of generalised interpersonal trust that is intended to form the baseline for international comparisons, and is the highest priority for inclusion in any attempt to measure trust.
- Module A contains those questions for which evidence of validity and policy relevance is the strongest and which are most apt to achieve some degree of international harmonisation. Unlike all of the other question modules included in these Guidelines, the core module is intended to be used without significant amendment and in full.
- Modules B to E are focused on different approaches to measuring trust. These modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires.

Survey Implementation

- Interviewer training is crucial to the quality of responses in any survey. To manage risks around respondent attitudes to questions on trust, interviewers should be well-briefed, not just on what concepts the questions are trying to measure, but also on how the information collected will be used.
- Evidence suggests that measures of trust are relatively non-problematic for respondents to answer. Item-specific non-response rates for interpersonal measures of trust are

similar to those for marital status, education and labour market status, and much lower than for those on income. Non-response rates for questions on institutional trust are somewhat higher, but still lower than is the case of income. This suggests that, in general, trust questions are not perceived as problematic by respondents.

- Normal data-cleaning procedures include looking for obvious errors such as transposed numbers, duplicate records, loss of records, incomplete responses, out-of-range responses or failure to follow correct skip patterns. Some issues are of particular relevance to trust data. In particular, where a module comprising several questions with the same scale is used, data cleaning should also involve checking for response sets (see Chapter 3).

Notes

1. The distinction between cardinal and ordinal measures is important when measuring trust. With ordinal measures, the responses are assumed to show the rank-order of different states, but not the magnitude. For example, with ordinal data a 5 is considered higher than a 4, and an 8 is considered higher than a 7; however, nothing can be said about the relative size of the differences implied by different responses. Conversely, for cardinal data it is assumed that the absolute magnitude of the response is meaningful, and that each step on the scale of responses represents the same amount. Thus, a person reporting a level of trust of 5 would deem strangers to be more trustworthy than someone reporting a 4 by the same amount as someone reporting an 8 compared to a 7. There is little direct evidence on how trust measures should be treated. In principle, such measures are ordinal, but it is worth noting that in the case of subjective well-being – which presents a similar measurement challenge – there is strong evidence suggesting that treating the measures as cardinal does not bias the results obtained (Ferrer-i-Carbonell and Frijters, 2004).
2. Equivalised income is a measure of the level of economic resources available to the 'consumption units' in a household. Total household income is adjusted by the size of the household to reflect the fact that larger households will have to spread a given level of income more widely over the members, but also that some economies of scale in consumption are possible. A commonly used approach to equivalising income is to divide household income by the square root of the number of people living in the household.
3. While many surveys provide information on whether respondents live in an urban or rural area, the use of such data raises issues for international comparability. This is because, in some cases, respondents may self-report whether they live in an urban or rural area, while in others they are classified as urban or rural according to their socio-economic conditions, and in other cases yet the urban/rural classification may rely on a settlement-based approach. Because of these differences in national practices, a better solution could be for interviewers to classify respondents as urban/rural based on information on the characteristics of their place of living. OECD TL3 regions are classified as *predominantly urban*, *predominantly rural* or *intermediate* based on the population density within the region, based on the same method and thresholds across countries; in the case of the larger OECD TL2 regions, the classification is based on the share of population living in each functional urban area.
4. Where these detailed labour force guidelines cannot be implemented in full because of space constraints, employment status can be reported based on the self-assessment of respondents.
5. The OECD Guidelines on Measuring Subjective Well-being (2013) provide advice on the use of subjective well-being measures in a range of fields.
6. This is not, in fact, beyond the realm of possibility. Many government agencies may have an interest in collecting measures of clients' satisfaction with the services that these agencies provide and in client trust of the agency.
7. Deaton (2011) finds a large effect for Valentine's Day on measures of subjective well-being in the United States, despite the day not being a public or bank holiday.
8. Internet surveys are, from this perspective, a way of implementing CASI.
9. In fact, the OECD Guidelines on Measuring Subjective Well-being (2013) recommend that the core questions on subjective well-being are placed at the start of relevant surveys. For this reason, these Guidelines do not make the same recommendation for measures of trust.
10. In this case the precise transition question used was: "Now thinking about your personal life, are you satisfied with your personal life today?", and the subjective well-being measure that followed

was the Cantril self-anchoring ladder of life measure. It does not follow that the same transition question will work in other contexts, and transition questions should be tested empirically before being relied on.

11. The Rosenberg question is: "Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people?" (Rosenberg, 1957).

12. See Annex 1 for the wording of the lost wallet question.

13. One key difference between the questions used here and Glaeser's original questions relates to the use of the phrase *your rooming group's hallway door* which is replaced by the phrase *your door* in the question used here. This reflects the fact that Glaeser's questionnaire was intended for students living on campus at an American university.

References

Alesina, A. and E. La Ferrara (2001), "Who trusts others?", *CEPR Discussion Paper,* No. 2646.

Algan, Y. and P. Cahuc (2013), "Trust, growth and well-being: New evidence and policy implications", *IZA Discussion Paper,* No. 7464.

Algan, Y. and P. Cahuc (2010), "Inherited trust and growth", *The American Economic Review*, Vol. 100, No. 5, pp. 2060-2092.

Berg, J., J. Dickhaut and K. McCabe (1995), "Trust, reciprocity, and social history", *Games and Economic Behavior*, Vol. 10, pp. 122-142.

Boarini, R. et al. (2013), "What makes for a better life? The determinants of subjective well-being in OECD countries – Evidence from the Gallup World Poll", *OECD Statistics Working Papers*, No. 2012/03, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k9b9ltjm937-en*.

Borgonovi, F. and T. Burns (2015), "The educational roots of trust", *OECD Education Working Papers,* No. 119, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5js1kv85dfvd-en*.

Bouckaert, G. (2012), "Trust and public administration", *Administration*, Vol. 60, No. 1, pp. 91-115.

Bouckaert, G. and S. Van de Walle (2003), "Comparing measures of citizen trust and user satisfaction as indicators of 'good governance': Difficulties in linking trust and satisfaction indicators", *International Review of Administrative Sciences*, Vol. 69, No. 3, pp. 329-343.

Bycroft, C. (2011), *Social and Population Statistics Architecture for New Zealand*, Statistics New Zealand, Wellington.

Cox, J. (2004), "How to identify trust and reciprocity", *Games and Economic Behaviour*, Vol. 46, pp. 260-281.

Deaton, A. (2011), "The financial crisis and the well-being of Americans", *Oxford Economic Papers,* No. 64, pp 1-26.

Delhey, J., K. Newton and C. Welzel (2011), "How general is trust in 'most people'? Solving the radius of trust problem", *American Sociological Review*, Vol. 76, No. 5, pp. 786-807.

Dupré, D. and E. Di Meglio (2014), *Planned Future Developments of EU-SILC*, EU-SILC Conference, Lisbon.

Eurostat (2005), *Guidelines for the Development and Criteria for the Adoption of Health Survey Instruments*, Eurostat, Luxembourg.

Ferrer-i-Carbonell, A. and P. Frijters (2004), "How important is methodology for the estimates of the determinants of happiness?", *The Economic Journal*, No. 114, pp. 641-659.

Fleischer, L., C. Smith and C. Viac (2016), "A Review of General Social Surveys", OECD Statistics Working Papers, No. 2016/09, OECD Publishing, Paris, *http://dx.doi.org/10.1787/bb54d16f-en*.

Fukuyama, F. (1995), *Trust: The Social Virtues and the Creation of Prosperity*, Hamish Hamilton, London.

Ginn, J. and S. Arber (2004), "Gender and the relationship between social capital and health", A. Morgan and C. Swann (eds.), *Social Capital for Health: Issues of Definition, Measurement and Links to Health*, Health Development Agency, London.

Glaeser, E. et al. (2000), "Measuring trust", *The Quarterly Journal of Economics,* Vol. 115, No. 3, pp. 811-846.

Goldberg, D.P. et al. (1978), *Manual of the General Health Questionnaire*, Windsor, England, NFER Publishing.

Helliwell, J. et al. (2014), "Good governance and national well-being: What are the linkages?", *OECD Working Papers on Public Governance*, No. 25, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jxv9f651hvj-en*.

Helliwell, J. and S. Wang (2010), "Trust and well-being", *NBER Working Papers,* No. 15911, National Bureau of Economic Research.

Intawan, C. and S.P. Nicholson (2017), "My trust in government is implicit: Automatic trust in government and system support", unpublished manuscript.

Kahneman, D., J. Knetsch and R. Thaler (1986), "Fairness and the assumptions of economics", *The Journal of Business*, Vol. 59, No. 4, pp. S285-S300, *www.jstor.org/stable/2352761*.

McKnight, D.H., V. Choudhury and C. Kacmar (2002), "The impact of initial consumer trust on intentions to transact with a web site: A trust-building model", *The Journal of Strategic Information Systems*, Vol. 11, No. 3, pp. 297-323.

Naef, M. and J. Schupp (2009), "Measuring trust: Experiments and surveys in contrast and combination", *IZA Discussion Paper,* No. 4087.

Nannestad, P. (2008), "What have we learned about generalized trust, if anything?", *Annual Review of Political Science*, Vol. 11, pp. 413-436.

Nooteboom, B. (2007), "Social capital, institutions and trust", *Review of Social Economy,* Vol. 65, No. 1, pp. 29-53.

NORC (2014), *General Social Survey* (database), *https://gssdataexplorer.norc.org/*.

OECD (2017), *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264268920-en*.

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

ONS (2012), *Subjective Well-being: A Qualitative Investigation of Subjective Well-being Questions*, Office for National Statistics, UK, London.

Putnam, R. (1993), *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press, New Jersey.

Rosenberg, M. (1957), "Misanthropy and attitudes toward international affairs", *Journal of Conflict Resolution*, pp. 340-345.

Rothstein, B. and E. Uslaner (2005), "All for all: Equality, corruption and social trust", *World Politics,* Vol. 58, No. 3, pp. 41-72.

Schneider, I. (2016), "Can we trust measures of political trust? Assessing measurement equivalence in diverse regime types", *Social Indicators Research, http://dx.doi.org/10.1007/s11205-016-1400-8*.

Scrivens, K. and C. Smith (2013), "Four interpretations of social capital: An agenda for measurement", *OECD Statistics Working Papers*, No. 2013/06, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jzbcx010wmt-en*.

Shanahan, E. (2007), "Excuse me, is this your phone?", *Reader's Digest,* July.

Soroka, S., J. Helliwell and R. Johnston (2003), "Measuring and modelling trust", *Diversity, Social Capital and the Welfare State,* University of British Columbia Press, Vancouver.

Stafford, M. et al. (2005), "Gender differences in the associations between health and neighbourhood environment", *Social Science and Medicine,* Vol. 60, No. 8, pp. 1681-1692.

Statistics New Zealand (2015), *A matter of trust: Patterns of M ori trust in institutions 2013*, Wellington.

Stiglitz, J.E., A. Sen and J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, *www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf*.

UNECE (2013), *Survey Module for Measuring Health State,* United Nations, New York and Geneva.

UNECE (2010), *Manual on Victimisation Surveys*, United Nations, New York and Geneva.

United Nations (2014), *Fundamental Principles of Official Statistics*, A/RES/68/261, United Nations, New York and Geneva.

United Nations Statistical Division (1986), *National Household Survey Capability Programme, Sampling Frames and Sample Designs for Integrated Survey Programmes. Preliminary version*, United Nations, New York.

United Nations Statistical Division (1984), *Handbook of Household Surveys*, United Nations, New York.

Uslaner, E. (2008), "Trust as a moral value", D. Castiglione, J. Van Deth and G. Wolleb (eds.), *The Handbook of Social Capital*, Oxford University Press, Oxford, pp. 101-121.

Uslaner, E. (2002), *The Moral Foundations of Trust*, Cambridge University Press.

WHO (2012), *World Health Survey Instruments and Related Documents*, World Health Organisation, Geneva.

© OECD 2017

Chapter 5

Output and analysis of trust data

This chapter provides guidance for data producers, media and researchers on how to deal with trust data once they have been collected. The chapter discusses the planning of statistical releases of trust data for a range of target audiences and highlights practical examples of various reporting styles. Advice is provided on the interpretation of results and analysis of microdata, including their challenges.

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

5.1. Introduction

After having discussed validity and measurement and the logistics of data collection in Chapters 2 to 4, the present chapter closes the circle by considering *how to use* trust data once they have been gathered. In other terms, what is best practice in releasing trust data? What key interpretative issues should be kept in mind when examining the data? What basic analytical techniques should (and should not) be applied?

The chapter has three main sections. Section 5.2 starts with the reporting of trust data and highlights the different types of end users of statistical releases, whose needs and capacities data analysts at national statistical offices have to be aware of. This section focuses on the type and depth of information that is most appropriate for different audiences and, to showcase best practice, provides examples of how national statistical offices (NSOs) and other data producers have released their results.

The chapter then turns to the interpretation and analysis of trust data. Section 5.3 discusses how to evaluate differences in trust between observations (e.g. groups of people or countries, and over time) and how to determine what to consider as *big* rather than *small* differences.

Section 5.4 discusses the empirical analysis of trust data – considering trust both as a driver of other outcomes of interest (e.g. economic growth, subjective well-being, health-seeking behaviour) and as a valuable outcome in its own right, for which we want to identify drivers (e.g. socio-demographic factors, life experiences, experiences with government). Basic analytic methods that are appropriate for selected cases are also discussed.

Section 5.2, on reporting, will be of most direct interest to statistical analysts working for producers of large-scale data, such as NSOs, as it concerns the kinds of outputs and analyses that they are most likely to report for different types of audiences. Section 5.3., dealing with interpretation, and Section 5.4, on analysis, are relevant to the broader research and policy audience concerned with trust.

Generally, this chapter's messages apply equally to measures of trust in both other people and institutions. Therefore, unless otherwise specified, the term trust as used in this chapter refers to both types of trust.

5.2. Reporting trust data

Ideally, NSOs would regularly collect and release high-quality trust data from large and representative population samples. As for other outcome measures, key users of trust data include policy makers, public service providers, civil society organisations, researchers and the wider public – all of whom may have an interest in whether, where and when conditions in society are improving or deteriorating. Given the heterogeneity of these audiences in terms of their prior knowledge and quantitative literacy, data producers should ensure that their releases mean something to the general public as well as to specialists (New Economics Foundation, 2009).

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

Statistical analysts in NSOs therefore have an important communication role, as they are the ones responsible for tailoring and presenting statistical information in the most appropriate way for these multiple audiences. Statistical releases can be thought of in terms of having different layers, or pages (see Table 5.1 for an illustration). The ***front page*** will mainly provide the general public as well as policy makers with a quick overview of the data. This front page should concisely yet precisely report levels of a single headline measure, or at most a small set of key measures, for strategic communication purposes and in order to avoid overloading laypeople with information. The core trust measures based on the prototype modules recommended by these Guidelines (see Annex 2) are ideally suited to feature on such a front page. Statistical staff should further include a brief commentary for the media, reiterating the importance and usefulness of measuring trust and describing the indicators used, the questions they are based upon, any chosen reporting thresholds, as well as the sample size and response rates. The importance of such commentary cannot be overstated: it immensely increases the likelihood that statistical releases are taken up by media and, given that news outlets often reprint the statistical analyst's commentary directly, they will also ensure that key features of the data are accurately reported. This is all the more important since very often the general public and policy makers will not read the statistical release itself, but are more likely to rely on a variety of media reports, including newspaper articles or social media posts.

Table 5.1. **Suggested structure of statistical releases of trust data**

	Purpose	Target audience	Specific content	General content
Front page	Give concise and precise overview of current trust levels	• Policy makers interested in a quick overview- General public • Most media outlets	• Headline measure levels • Commentary for media	All pages should include full descriptions of the question wording, answering scales, the sample size and any chosen thresholds
Second page	Provide more detailed information (e.g. trends over time, distributions across different population groups)	• Government analysts • Media outlets interested in a deeper analysis of trust	• Additional trust measure levels • Breakdown by socio-economic groups (including standard deviations) • Previous trust measure levels	
Third page	Facilitate access to and use of microdata	• Researchers and analysts wanting to conduct further analysis	• Microdataset • Questionnaire	

StatLink http://dx.doi.org/10.1787/888933584241

The ***second and third pages*** of statistical releases cover trust data in greater detail. The second page usually targets policy analysts, journalists and researchers who want to analyse trust beyond singular "quick and dirty" headlines. This analysis might look at trends in trust and its distribution, i.e. how trust has changed over time, and whether trust varies between groups (from different demographic groups to regions within a country up to the international level). This page hence expands the front-page content by reporting on relevant trust items other than the core headline measures (i.e. the prototype question Modules B-F described in Annex 2). Data can be further broken down by different population groups (e.g. by age, gender, employment status, income quartile, marital status, region of residence, and other factors identified as relevant for the specific country context), along with information on sample size, standard deviations, and the results of any significance tests for group differences that have been conducted. If available, trust data for earlier points in time should also be included for comparison purposes. In case the question wording, response scale or survey vehicle have changed over time, this should be pointed out.

The third page of data releases is designed for researchers and policy analysts who want to conduct deeper analysis with the trust data. In addition to directions on where and how to

access the trust microdata (ideally microdata should always be made publicly available), this page features a full description of the dataset and a copy of the full questionnaire, so that researchers can assess the order in which trust items were placed within the wider survey and take joint distributions with other relevant variables into account.

In general, audiences will find data releases easier to digest if there is some degree of consistency between the way the latest trust levels are reported on the first page of statistical releases and the way more detailed trends are described on the second page. Different ways of presenting and visualising trust data in practice are illustrated in Table 5.1.

The front page of statistical releases: Reporting levels of trust

The level of a trust measure essentially answers the question of whether the amount of trust in the population of interest is high or low. There are three main approaches to report that level. First, the frequency of responses can be described by category, i.e. the proportion of the population who select each response category of the respective response scale. The other approaches summarise the data for the purpose of easier presentation: this can be done either in relation to one or more threshold (e.g. the proportion of the population falling above or below the chosen threshold) or via a measure of central tendency (e.g. the mean, the median or the mode).[1]

The section below outlines the advantages and disadvantages of each of these three approaches, with visual examples of relevant statistical releases by either NSOs or other data producers. While data producers could refer to this for general advice on what type of reporting is more or less useful for trust measures, the final choice on presentation will also depend on the exact purpose of the release at hand.

Reporting results by frequency in each category

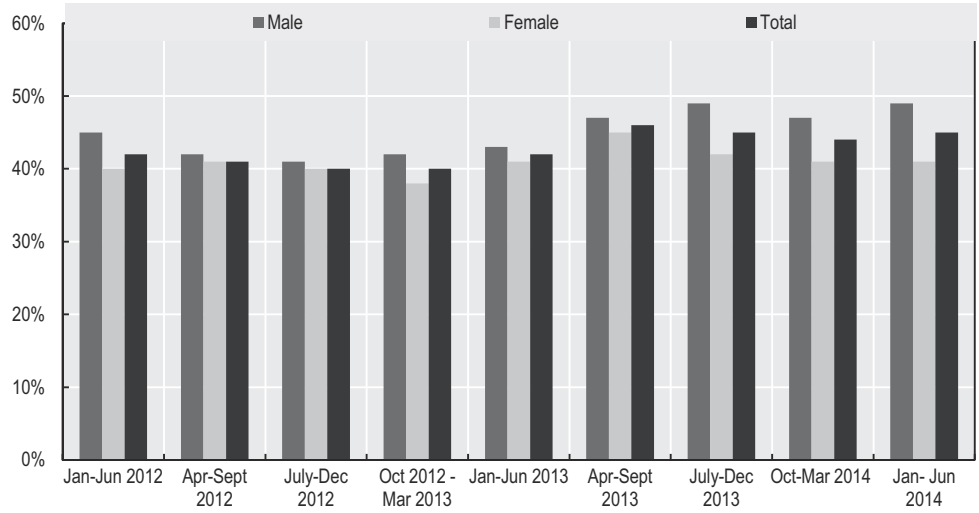
At first glance, indicating the proportion of respondents selecting each response category seems appealing: as no categories are merged and the entire distribution is described, no information is lost, and the data producer makes no (arbitrary) decisions about how to collapse and present data. Nevertheless, there are good reasons why this type of presentation might not be appropriate for trust data. Unless the number of categories is limited, presenting the whole distribution of responses for each measure is likely to be overwhelming: non-technical audiences will, without some guidance about the *story* told by the data, not be able to directly compare and evaluate such distributions. Furthermore, while this approach works reasonably well for dichotomous trust measures (where reporting the share of responses having chosen to either trust or distrust is equivalent to reporting each response category), Chapter 3 of the Guidelines has provided arguments as to why a 0-10 response scale is preferable over a dichotomous one. Since a 0-10 response scale is relatively long, reporting the share of respondents for each of the 11 points on the scale is not recommended.

Reporting results by proportions above or below thresholds

Reporting on the proportion of responses falling above or below one or a set of thresholds offers a way around the problems of managing a large number of scale responses and facilitating respondent comprehension of the story behind the data. For example, responses can be either reported as the percentage of respondents falling above or below a certain cut-off point or collapsed into *high*, *medium* and *low* categories. Figures 5.1 and 5.2 below provide examples of each approach.

OECD GUIDELINES ON MEASURING TRUST © OECD 2017

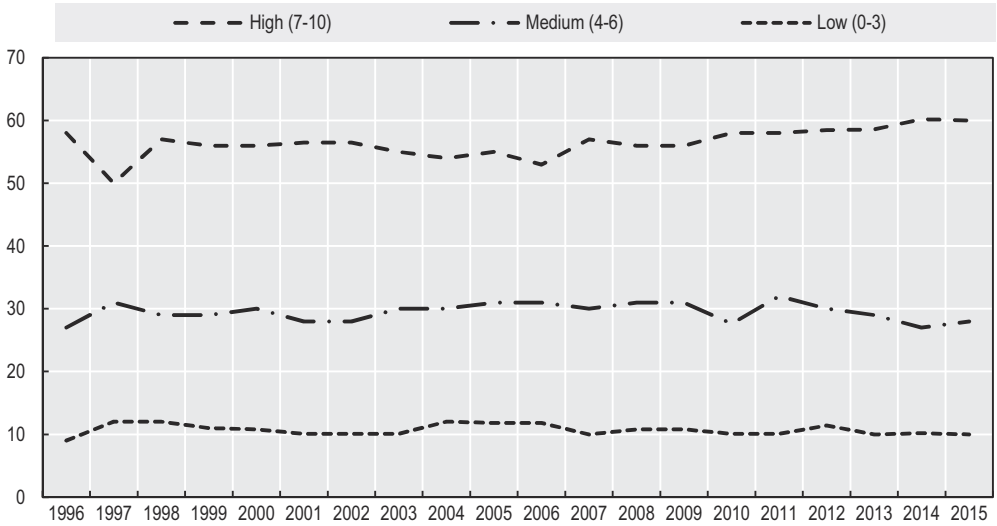
Figure 5.1. **Trust in public service by gender in New Zealand**



Note: Response options range from 1 ("Do not trust them at all") to 5 ("Trust them completely") to the question: "Overall, to what extent do you trust the public service?" The data show the percentage of respondents who gave a 4 or 5 out of 5.
Source: New Zealand State Services Commission (2014), Kiwis Count Survey (database), *www.ssc.govt.nz/kiwis-count-datasets*.

StatLink *http://dx.doi.org/10.1787/888933583880*

Figure 5.2. **Interpersonal trust in Sweden, 1996-2015**



Note: Response options range from 0 ("It is not possible to trust people in general") to 10 ("It is possible to trust people in general") to the question: "According to your view, to what extent is it possible to trust people in general?"
Source: SOM Institute (2015), The National SOM Survey (database), *http://som.gu.se/som_institute/-surveys/national-som*.

StatLink *http://dx.doi.org/10.1787/888933583899*

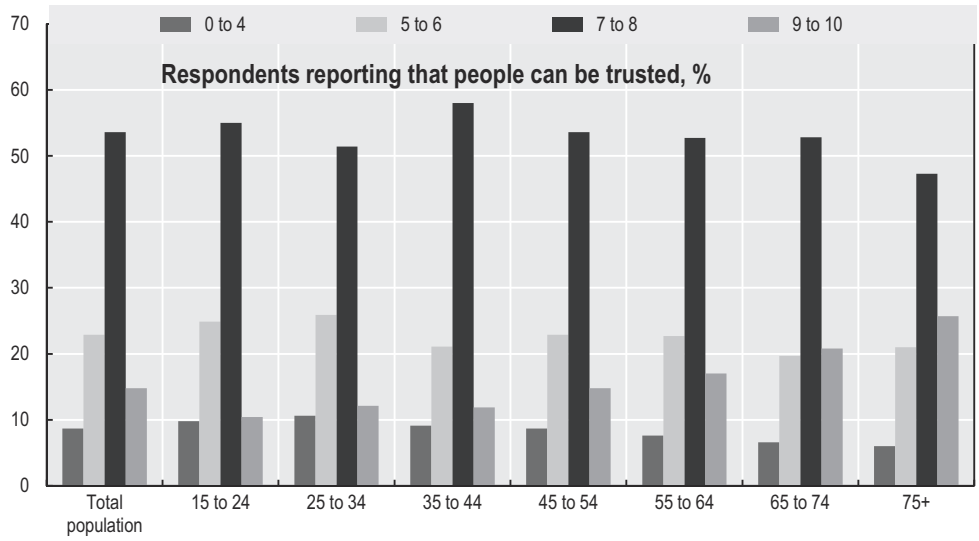
Figure 5.1 shows the proportion of the New Zealand population reporting trust in public services from 2012 until 2014, disaggregated by gender. The threshold for trust in this case was selected to be "4 and above" on a 5-point scale (where 1 is *do not trust them at all* and 5 is *trust them completely*) when answering the question: *Overall, to what extent do you trust the public service?* Figure 5.2 provides an example of collapsing data into several threshold-based categories. The data (from the SOM institute, an independent survey research organisation at the University of Gothenburg in Sweden) refer to responses to the question, "according to

your view, to what extent is it possible to trust people in general?", based on a response scale from 0 ("not possible") to 10 ("possible"). The SOM researchers coded respondents selecting the response categories 0-3 as having *low* trust in people in general, those ticking categories 4-6 as having *medium* trust, and those ticking 7-10 as having *high* trust. The suggested output for the question modules presented in Annex 2 also introduces the possibility of reporting the share of responses falling *below* the threshold of 4 on a 0-10 scale, hence focusing on the part of the population that does not trust other people or institutions.

Both figures make the benefits of threshold-based measures salient – the threshold provides an anchor and reference frame for interpretation, and thus eases the digestion of the results. However, threshold-based measures are far from perfect. Any summary measure, by definition, can be misleading, as it does not provide information on the distribution of data. Collapsing categories means losing some of the data richness and can lead to a narrow policy focus on shifting people from just below to just above a threshold. This is particularly problematic if only one threshold at the upper end of the response scale is used (as, for example, in Figure 5.1), as it may be important, for policy makers in particular, to recognise the existence of people at both extremes of the trust spectrum. In addition, reporting estimates based on thresholds runs the risk of presenting two very similar distributions as very different, when the only differences are around the threshold, or vice versa (OECD, 2013). Lastly, the use of threshold-based measures risks arbitrarily assigning individuals to high, medium and low trust categories without much evidence of what these categories mean in practice, and without testing them for real-world validity (Blanton and Jaccard, 2006).

One compromise between managing scale length by collapsing response scales and not assigning arbitrary and potentially misleading categories is to avoid labelling the categories as high or low, but to neutrally refer to their place on the scale, for example "0-4" or "5-6" (see Figure 5.3 for an example). However, even with this approach, it is unclear which criteria govern the selection of the scale ranges to be bundled together. Overall, given the challenges

Figure 5.3. **Interpersonal trust by age groups in New Zealand, 2014**



Note: Based on a scale where 0 is "not trusted at all" and 10 is "trusted completely" as answer to the question: "In general how much do you trust most people in New Zealand?"
Source: Stats New Zealand (2014), New Zealand General Social Survey (database), *www.stats.govt.nz/nzgss2014*.

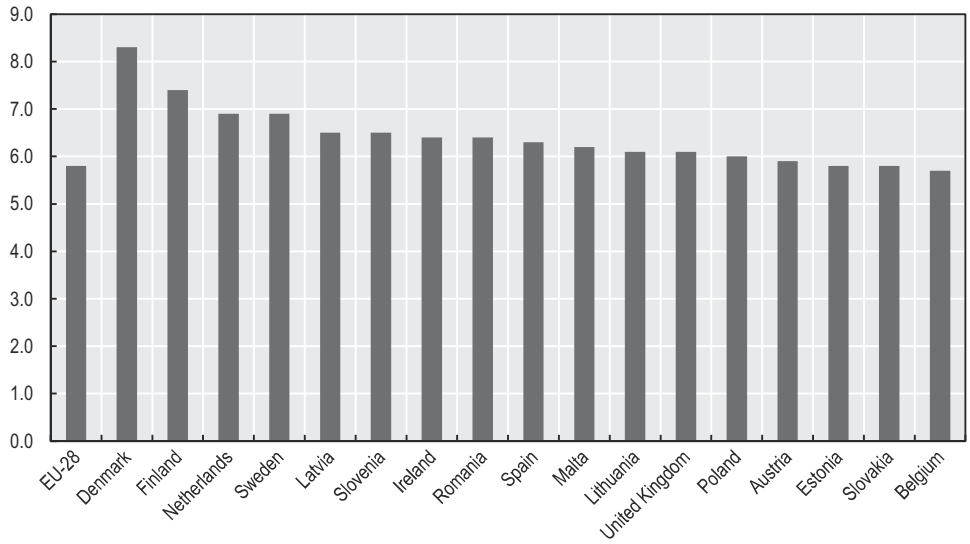
StatLink *http://dx.doi.org/10.1787/888933583918*

associated with setting suitable thresholds, statistical analysts should avoid adopting these types of summary measures uncritically. At the very least, information about which threshold was chosen should always be presented alongside such data.

Reporting results by central tendency measures

The third approach for presenting levels of trust involves summarising the central tendency of a distribution in a single and easily understandable number. While the median and the mode may lack sensitivity to changes over time or to differences between groups due to the limited number of scale categories, the mean is a more useful summary statistic. Figure 5.4 shows the mean average value (on a 0-10 response scale) to EU-SILC's generalised trust question, whereas Figure 5.5 does the same for the measure on trust in public institutions from the same survey. Figure 5.5 also highlights another good practice when reporting data on trust in institutions, i.e. separately showing results for various institutions (in this case, the political system, the police and the legal system) rather than merging them into some sort of summary index. Since the mean, as all other measures of central tendency, consists merely of a single number that takes up little space in data releases, this leaves more room for comparisons between different institutions in the same figure or table.

Figure 5.4. **Interpersonal trust in European countries, 2013**



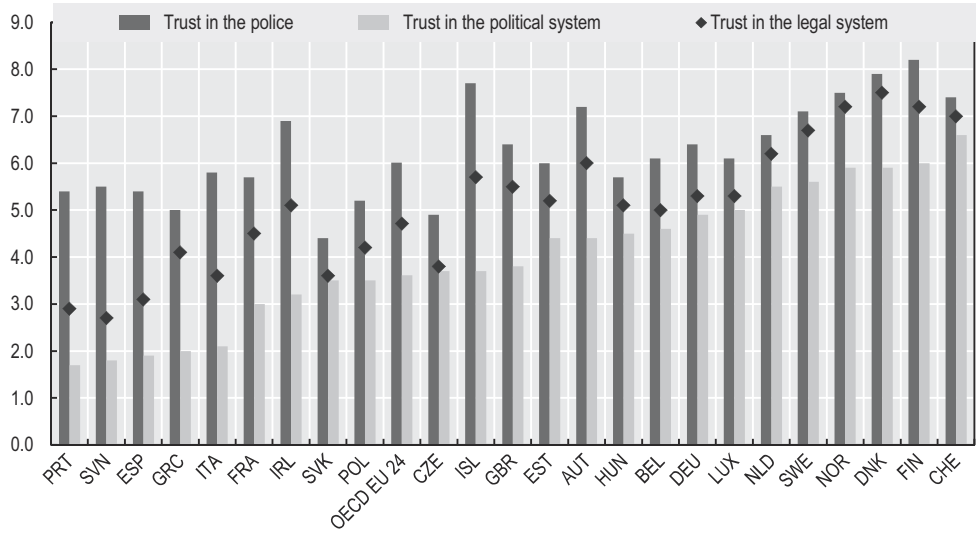
Note: Response options ranged from 0 ("You do not trust any other person") to 10 ("Most people can be trusted") to the question: "Some people say that you can trust most people. Others think you cannot be too careful in dealing with other people. Do you think most people can be trusted?" The OECD EU average is the population-weighted average of the values included in the chart.

Source: Eurostat (2015), European Union Statistics on Income and Living Conditions (EU SILC) (database), *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en*.

StatLink *http://dx.doi.org/10.1787/888933583937*

Of course, the mean has disadvantages as well: in particular, it requires treating data as cardinal even when they are, in fact, ordinal.[2] However, several studies suggest that this does not generally lead to biased results (Diener and Tov, 2012). More importantly, though, outliers can strongly affect the mean value, and the mean does not provide information about the data's underlying distribution. Therefore, the mean should be complemented with

Figure 5.5. **Trust in public institutions in European countries, 2013**



Note: Response options range from 0 ("No trust at all") to 10 ("Complete trust") to the question: "How much do you trust: The political system in [country]/ The legal system in [country]/ The police in [country]?" The OECD EU average is the population-weighted average of the values included in the chart.
Source: Eurostat (2015), European Union Statistics on Income and Living Conditions (EU SILC) (database), *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en*.

StatLink *http://dx.doi.org/10.1787/888933583956*

information about the distribution of data. One obvious choice is the standard deviation, but this is quite abstract and not easy to communicate to a non-technical audience. Other measures of dispersion, such as interquartile ranges, can thus be considered as alternative, depending on the statistical literacy of the release's target population.

The second page of statistical releases: Reporting change over time and differences between groups

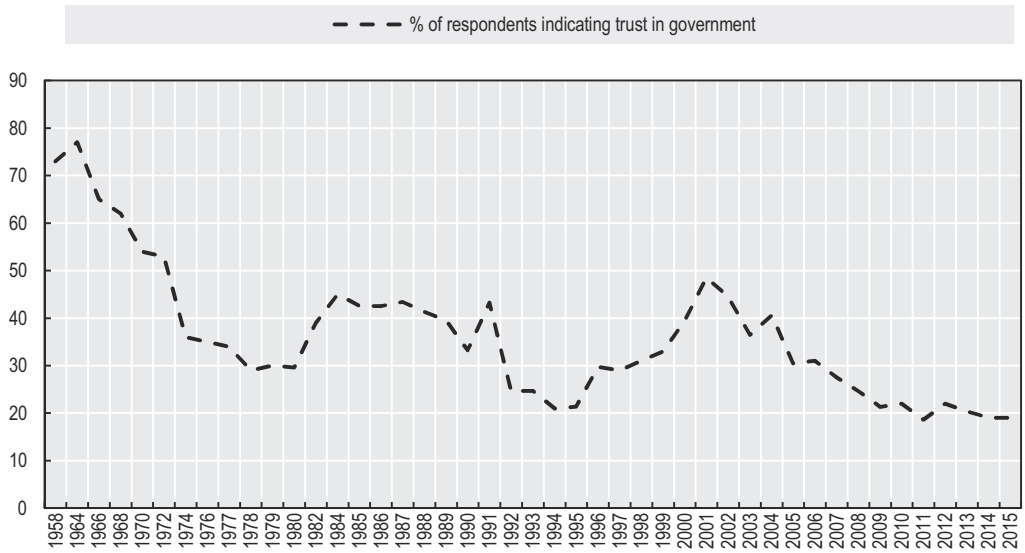
The report of trust data on the second page of statistical releases should generally cover two elements: changes over time, and differences between either societal groups or countries. Both of these elements are essential to provide some kind of "external reference point" to the headline trust measure of a country. How does the current level of trust compare to the score recorded one, two, three, four, five years ago? How does it compare to the trust score of other countries at similar levels of development? Do all groups within society have a similar score, or are there big discrepancies? More comprehensive reporting of trust measures should, ideally, answer all these questions.

In fact, all these questions point to the core mission of national statistics. If policy makers, civil society and researchers want to understand how to increase trust or to prevent its decay, the characteristics of the groups at both extremes of the trust scale need to be closely examined and better understood. Breaking down national statistics (e.g. by age, gender, education, place of residence, ethnicity, religion, occupation, socio-economic conditions, employment status, health status, etc.) and contrasting these measures of group performance against the overall population can enhance that understanding. Comparisons over time and between groups can also guide the quest for the potential drivers of trust.

Presenting measures of changes over time and of differences between groups requires reporting tools that are similar to those used for presenting central tendency measures. For example, changes over time can be monitored by tracking changes in mean trust through

time series (Figure 5.6) or by calculating changes in the mean score between various points in time (Figure 5.7).

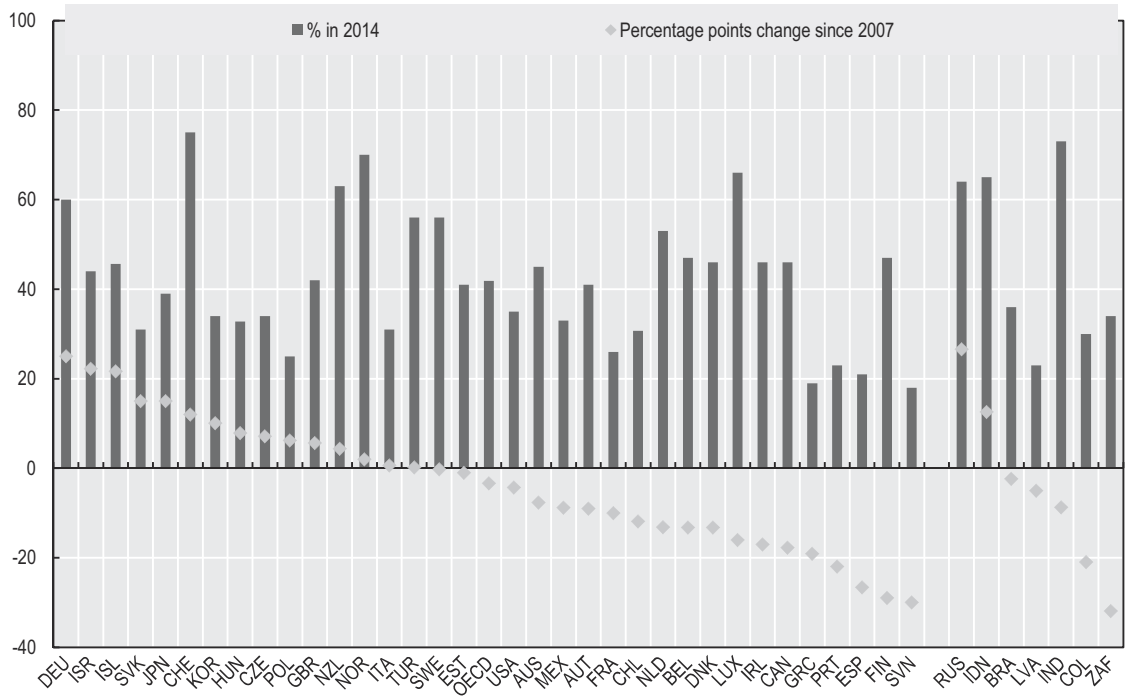
Figure 5.6. **Trust in government in the United States, 1958-2015**



Source: Pew Research Center (2016), *Beyond Distrust: How Americans View their Government, www.people-press.org/2015/11/23/1-trust-in-government-1958-2015.*

StatLink http://dx.doi.org/10.1787/888933583975

Figure 5.7. **Trust in national government in 2014 and changes since 2007, OECD countries**



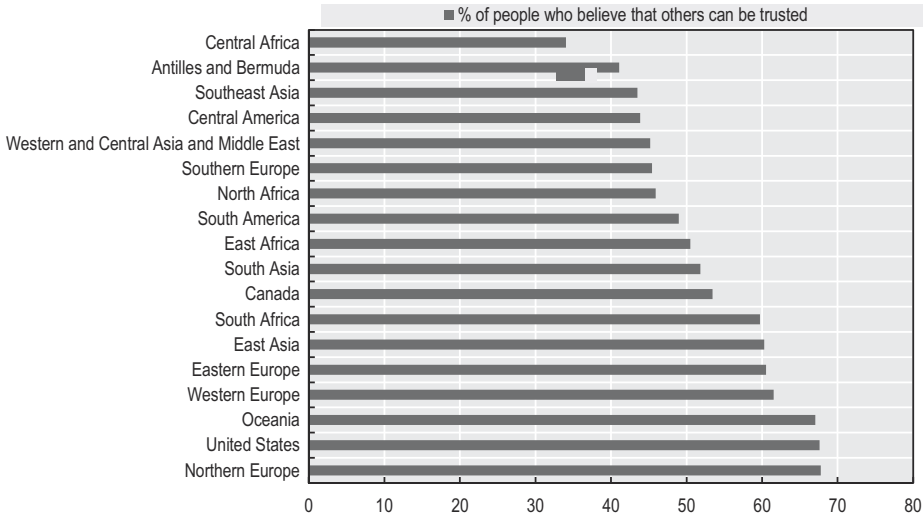
Note: Data refer to the percentage of people who answered *yes* to the question: "Do you have confidence in national government?" Countries are ranked in descending order according to the percentage point change between 2007 and 2014. Data for Chile, Hungary and Iceland refer to 2013 rather than 2014. Information on data for Israel: *http://dx.doi.org/10.1787/888932315602*.

Source: Gallup (2014), Gallup World Poll (database), *www.gallup.com/services/170945/world-poll.aspx*.

StatLink http://dx.doi.org/10.1787/888933583994

Group differences can be examined by presenting these differences over time, relative to a given threshold, or by showing the (absolute or percentage) differences in the proportion of respondents who have selected a specific answer for several groups of interest (Figure 5.8). Table 5.2 provides an example of how to include estimates of statistical inference testing.

Figure 5.8. **Interpersonal trust by region of birth in Canada, 2013**



Source: Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm.*

StatLink *http://dx.doi.org/10.1787/888933584013*

Table 5.2. **Interpersonal trust by population group in Canada in selected years**

People who reported most people can be trusted[1]

	2003	2008	2013
		percentage	
Total	55	48†	54†
Men (ref.)	57	48†	55†
Women	54*	47†	52*†
Age group			
15 to 24 years	52*	47†	48*†
25 to 34 years	51*	45†	52
35 to 44 years (ref.)	56	47†	53†
45 to 54 years	61*	50*†	55†
55 to 64 years	58	49†	57*
65 to 74 years	53	47†	55
75 years and older	54	48†	54
Province			
Newfoundland and Labrador	64*	49†	56†
Prince Edward Island	69*	51†	63
Nova Scotia	62	53†	59
New Brunswick	56*	47†	51*†
Quebec	35*	32*†	36*
Ontario (ref.)	60	51†	57†
Manitoba	64*	52†	58†
Saskatchewan	67*	54†	60†
Alberta	63*	55*†	60*
British Columbia	65*	57*†	63

1. "Generally speaking, would you say that most people can be trusted or that you cannot be too careful with people?"
† Significantly different from 2003 ($p < 0.05$).
* Significantly different from the reference category ($p < 0.05$).
Source: Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm.*

StatLink *http://dx.doi.org/10.1787/888933584260*

Depending on the target audience's interest, other relevant changes in trust (e.g. changes in the overall distribution, or differences in rates of change between different societal groups) can additionally be reported.

Since both sample size and standard errors are essential when comparing two or more observations, this information should be reported alongside group means (OECD, 2013). This information is easier to interpret when presented graphically, e.g. though box plots or error bars.

5.3. Interpreting trust data

This section focuses on how to interpret trust data after they have been released. Apart from NSO staff, this information is relevant for professional communities (i.e. media, think tanks) concerned with communicating the key messages and trends of fresh data releases, as well as for researchers and analysts working directly with microdata. Most analyses of trust data will naturally be concerned with examining *differences* between observations. Two essential questions to answer for these types of analyses will be whether these differences are actually meaningful in practical, real-life terms and to what extent any differences are due to measurement artefacts and errors. Both issues are addressed below.

Essentially, any assessment of the meaningfulness of differences between observations requires an understanding of what size of difference is likely to be encountered, and what difference can be categorised as *small* or *big*. These evaluations, by their nature, will be relative – for example, in comparison to values from other groups, countries or previous time points. A necessary precondition for making such relative comparisons is the availability of all relevant comparison points – in other words, knowledge about the full data universe of trust measures.

At present, knowledge about the trust data universe, its properties and boundaries is incomplete – which is in fact one of the *raisons d'être* for these Guidelines. Nevertheless, Tables 5.3 to 5.6 represent a first attempt to summarise what is currently known about differences between observations for trust measures. For measures of interpersonal trust and trust in institutions (the police, judicial system and government), each table details the size of the differences for three types of measure (i.e. a 0-10 scale, a threshold measure of 7 and above for 0-10 scales, and a binary measure) across the two dimensions of comparison (cross-sectional or between observations, and over time or within observations).

Within the cross-section dimension, differences are further broken down by differences between groups (age groups, groups with different educational attainments and different labour force status were selected as examples, as these are common socio-economic and demographic distinctions) and differences between countries. The size of the difference is presented relative to the respective sample mean – for example, the first cell in the first row of Table 5.3 indicates that, compared to the overall 2013 EU-SILC mean score on a 0-10 scale, the respondents aged 16 to 24 reported a level of interpersonal trust 0.2 scale points higher. The same column, for the threshold measure, shows that, compared to the total sample percentage of the 2015 United Kingdom ONS National Opinion Survey, 16.7% fewer respondents aged 16 to 44 chose an interpersonal trust score of 7 or above on a 0-10 scale (ONS, 2016). For differences between countries, the range of differences from the respective sample mean is shown; for instance, in the case of 0-10 scale trust measures, the score of the country with the lowest interpersonal trust (Bulgaria) was 1.6 points below the EU-SILC 2013 sample mean, while that of the country with the highest score (Denmark) was 2.5 points above the sample mean, resulting in an overall range of 4.1 points.

Table 5.3. **Magnitude of differences for measures of interpersonal trust**

	Cross-section					Over time					
	Between population groups			Between countries		Within individuals			Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source	Difference in mean scores		Source
0-10 scale	By age By employment status By level of education	16-24 years: 0.2 25-49 years: 0 50+ years: 0.06 In employment: 0.1 Unemployed: -0.5 Tetriary education: 0.5 Upper sec. education: -0.1 Lower sec. education -.0.2	EU-SILC (2013)	Low end of distribution: -1.6 (Bulgaria) High end of distribution: 2.5 (Denmark)	EU-SILC (2013)	n/a		n/a	2004-06 2004-10 2004-14	Minimum: 0.02 (Switzerland) Sample average: |0.2| Maximum: -0.56 (Ireland) Minimum: -0.012 (Finland) Sample average: |0.25| Maximum: 0.78 (Poland) Minimum: -0.01 (Norway) Sample average: |0.23| Maximum: -0.75 (Ireland)	European Social Survey
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	% of respondents that switched 7-10 threshold category		Source	Difference in % of respondents that selected 7-10		Source
	By age By employment status By level of education	16-44 years:-16.7% 45+ years: 6.5% In employment: 3.6% Unemployed: -0.7 % With degree/equivalent: 12.1% Below degree level: 1.5% No qualification: -4.6%	ONS National Opinion's Survey (2015)	Low end of distribution: -21.2% (Poland) High end of distribution 31.7% (Denmark)	European Social Survey (2014)	n/a		n/a	2004-06 2004-10 2004-14	Minimum: -0.05% (Ukraine) Sample average: |3.47%| Maximum: -10.65% (Ireland) Minimum: -0.07% (Spain) Sample average: |4%| Maximum: -13.54% (Ireland) Minimum: -0.22% (Norway) Sample average: |3.71%| Maximum: -11.86% (Ireland)	European Social Survey
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	% of respondents that switched binary category		Source	Difference in % of respondents that selected "trust"		Source
	By age By employment status By level of education	16-44 years: -0.3% 45+ years: 0.2% In employment: 2.1% Unemployment: -3.5% With degree/equivalent: 19% Below degree level: -3.7% No qualification: -12.7%	ONS National Opinion's Survey (2015)	Low end of distribution: -21.1% (Phillipines) High end of distribution: 43.5% (Netherlands)	World Values Survey (Wave 6, 2010-14)	1998-2000 2003-05 2007-08 1998-2006	23.69% 24.07% 22.77% 52.59%	British Household Panel Study	Wave 5 (2005-08) – Wave 6 (2010-14)	Minimum: -0.29% (Malaysia) Sample average: |5.64%| Maximum: 22.94% (Netherlands)	World Values Survey

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584279

Table 5.4. **Magnitude of differences for measures of trust in the police**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
0-10 scale	By age	16-24 years: -0.2 25-49 years: -0.3 50+ years: 0.17	EU-SILC (2013)	Low end of distribution: -2.4 (Bulgaria) High end of distribution: 2.2 (Finland)	EU-SILC (2013)	2004-06	Minimum: -0.007 (France) Sample average: |0.19| Maximum: -0.69 (Ukraine)	European Social Survey
	By employment status	In employment: -0.03 Unemployed: -0.9				2004-10	Minimum: 0.07 (Norway) Sample average: |0.36| Maximum: -1.45 (Greece)	
	By level of education	Tetriary education: 0.3 Upper sec. education: -0.1 Lower sec. education: -0.2				2004-14	Minimum: -0.06 (Finland) Sample average: |0.43| Maximum: 1.52 (Czech Republic)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -25.79% (Poland) High end of distribution: 26.38% (Finland)	European Social Survey (2014)	2004-06	Minimum: -0.16% (United Kingdom) Sample average: |3.19%| Maximum: -8.22% (Ireland)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.09% (Hungary) Sample average: |5.6%| Maximum: -19.77% (Greece)	
	By level of education	n/a				2004-14	Minimum: -0.65% (Finland) Sample average: |7.53%| Maximum: 22.13% (Czech Republic)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -43.74% (Venezuela) High end of distribution: 28.26% (Niger)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Cameroon, Japan, Tanzania, Latvia, Canada) Sample average: |6.11%| Maximum: 25% (Chad)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Vietnam, Hungary, Bolivia, Germany) Sample average: |7.16%| Maximum: 32% (Nepal)	
	By level of education	n/a				2006-15	Minimum: 0% (Paraguay, Serbia, Canada, Kosovo, Belgium, Singapore, Botswana) Sample average: |8.96%| Maximum: 29% (Nepal)	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584298

Table 5.5. **Magnitude of differences for measures of trust in the judicial system**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
0-10 scale	By age	16-24 years: 0.3 25-49 years: 0.1 50+ years: 0	EU-SILC (2013)	Low end of distribution: -1.9 (Slovenia) High end of distribution: 2.9 (Denmark)	EU-SILC (2013)	2004-06	Minimum: -0.019 (Hungary) Sample average: |0.28| Maximum: -1.25 (Ukraine)	European Social Survey
	By employment status	In employment: 0.17 Unemployed: -1				2004-10	Minimum: 0.01 (Finland) Sample average: |0.46| Maximum: -1.63 (Greece)	
	By level of education	Tetriary education: 0.7 Upper sec. education: 0.1 Lower sec. education -0.6				2004-14	Minimum: 0.08 (Ireland) Sample average: |0.43| Maximum: 0.97 (Czech Republic)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -29.53% (Poland) High end of distribution: 31.92% (Denmark)	European Social Survey (2014)	2004-06	Minimum: 0.03% (Hungary) Sample average: |3.31%| Maximum: -10.19% (Ukraine)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.64% (Slovakia) Sample average: |6.09%| Maximum: -18.95% (Greece)	
	By level of education	n/a				2004-14	Minimum: -0.07% (Ireland) Sample average: |6.9%| Maximum: 16.36% (Norway)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -44.17% (Ukraine) High end of distribution: 37.83% (Singapore)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Kyrgyzstan, Ukraine, Japan, Mexico) Sample average: |6.33%| Maximum: 32% (Poland)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Canada, Zambia, Montenegro, Denmark, Greece) Sample average: |7.56%| Maximum: -29% (Uganda)	
	By level of education	n/a				2006-15	Minimum: 0% (Guatemala) Sample average: |9.68%| Maximum: -39% (Cyprus)[3, 4]	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584317

Table 5.6. **Magnitude of differences for measures of trust in government**

	Cross-section					Over time		
	Between population groups			Between countries		Within countries		
0-10 scale	Difference in mean scores (population group – total sample)		Source	Difference in mean scores (country – total sample)	Source	Difference in mean scores		Source
	By age	16-24 years: 0.5 25-49 years: -0.1 50+ years: 0.06	EU-SILC (2013)	Low end of distribution: -1.8 (Portugal) High end of distribution: 3.1 (Switzerland)	EU-SILC (2013)	2004-06	Minimum: -0.007 (Germany) Sample average: |0.28| Maximum: -1.44 (Ukraine)	European Social Survey
	By employment status	In employment: 0.17 Unemployed: -1.1				2004-10	Minimum: -0.05 (Portugal) Sample average: |0.61| Maximum: -2.29 (Greece)	
	By level of education	Tetriary education: 0.7 Upper sec. education: 0.2 Lower sec. education: -0.6				2004-14	Minimum: -0.1 (Belgium) Sample average: |0.52| Maximum: -1.19 (Slovenia)	
Threshold measure: 0-10 scale	Difference in % of respondents that selected 7-10 (population group – total sample)		Source	Difference in % of respondents that selected 7-10 (country – total sample)	Source	Difference in % of respondents that selected 7-10		Source
	By age	n/a		Low end of distribution: -12.86% (Slovenia) High end of distribution: 13.35 (Norway)	European Social Survey (2014)	2004-06	Minimum: 0.19% (Hungary) Sample average: |1.8%| Maximum: -7.31% (Ukraine)	European Social Survey
	By employment status	n/a				2004-10	Minimum: -0.44% (France) Sample average: |4.88%| Maximum: -10.87% (Greece)	
	By level of education	n/a				2004-14	Minimum: 0.59% (Belgium) Sample average: |5.31%| Maximum: 14.99% (Norway)	
Threshold measure: binary scale	Difference in % of respondents that selected "trust" (population group – total sample)		Source	Difference in % of respondents that selected "trust" (country – total sample)	Source	Difference in % of respondents that selected "trust"		Source
	By age	n/a		Low end of distribution: -37.71% (Ukraine) High end of distribution: 45.29% (Singapore)	Gallup World Poll (2015)	2006-08	Minimum: 0% (Peru, Singapore, Philippines, Turkey, Denmark) Sample average: |10.84%| Maximum: 39% (Ecuador)	Gallup World Poll
	By employment status	n/a				2006-12	Minimum: 0% (Vietnam, South Korea, Kazakhstan, Israel, Panama) Sample average: |12.97%| Maximum: 52% (Ecuador)	
	By level of education	n/a				2006-15	Minimum: 0% (Japan, Singapore) Sample average: |13.51%| Maximum: -46% (Cyprus)[5]	

Source: OECD calculations based on the OECD Trust Database.

StatLink http://dx.doi.org/10.1787/888933584336

For changes over time, the tables present, if data are available, differences within individuals and within countries for various time ranges. For instance, in Table 5.3, longitudinal data from the British Household Panel Study from 1998-2008 show the share of respondents who switched between *no trust* and *trust* categories on a binary interpersonal trust measure. In the case of within-country differences, the minimum, sample average and maximum differences for various points in time are shown.

These tables show clearly that our current picture of the trust data universe is far from complete. The lack of data is particularly salient for individual-level panel data – only Table 5.3, the one focusing on interpersonal trust, features results, and only for a dichotomous measure. Furthermore, the underlying data in all tables varies greatly in quality. While EU-SILC data represent official statistical measures, especially for longitudinal data, non-official surveys such as the European Social Survey, the Gallup World Poll and the World Values Survey had to be relied upon. These surveys often suffer from small sample sizes and varying degrees of quality control. In some cases, national rather than cross-country data were used (i.e. the ONS National Opinion Survey and the British Household Panel Study). Furthermore, the tables are not significantly expanding the data universe beyond the geographic boundaries of Europe, with the exception of the Gallup World Poll and the World Values Survey. Lastly, not all measures in the table are based on the same question wording – for instance, the European Social Survey asks about trust in politicians, rather than in government. Thus, while these tables can serve as a useful reference point for people interested in situating effect sizes in a comparative context, firm conclusions should not be drawn until better, high frequency, global trust data are collected.

Furthermore, although the tables focus on mean differences in trust, inequality of trust within groups, across countries and over time are additional aspects to consider. Although standard errors of trust estimates are currently often not examined or reported in policy publications, they can provide relevant information about the state of trust and point to differences in the distributions between observations. A similar exercise featuring differences in trust distributions should thus be repeated in the future.

Keeping these limitations in mind, some preliminary trends are nevertheless detectable:

- Differences in interpersonal trust tend to be smaller than differences in institutional trust across the board. For example, the within-country average difference in 0-10 scale measures of trust over the 2004-14 period is 0.23 scale points for interpersonal trust, 0.43 scale points for trust in the police, 0.46 scale points for trust in the judicial system and 0.52 scale points for trust in government.
- When examining cross-sectional differences between societal groups, the employment status and educational attainment groups display larger differences than age groups.
- Individual-level interpersonal trust, at least on a binary measure, seems to be much more volatile over time than aggregate measures of the same construct would suggest. British Household Panel Study data exhibit a high level of instability in individual-level reports of trust, with nearly half of all sample members changing their trust response at least once over the 1998-2006 period of observation.[6]
- The choice of measures can make a difference: for similar points in time, the 0-10 scale measure, the "7 and above" threshold measure and the binary measure do not always result in the same countries being at the bottom and top ends of the difference distribution. For instance, focusing on the 2004-06 period of observation using European Social Survey data on trust in the police, France displays the smallest difference of -0.007 scale points and

Ukraine the largest difference of -0.69 scale points for a 0-10 scale measure, whereas the United Kingdom and Ireland take these places for the threshold measure, with differences in respondents falling over the threshold of -0.16% and -8.22%, respectively.

- Differences between countries are larger than differences within countries over time for all measures and types of trust considered.
- Effect sizes seem to grow larger as the sample considered increases in geographical coverage beyond Western nations. This suggests that as many countries as possible should be considered in order to fully comprehend the trust data universe and the volatility of trust measures in different contexts.

In conclusion, what constitutes a small or big difference in trust depends on the nature of the difference considered. It is important to consider the full range of potential drivers when interpreting results, including the influences stemming from measurement error described in the next section. The key message here is that knowledge on the overall data universe of trust will remain limited until higher quality and more frequent data become available. The current lack thereof significantly impacts the interpretations and conclusions that can be drawn about the magnitude of effect sizes.

Alternative influences on effect sizes and the possible impact of culture

The large differences between trust measures can be influenced by a variety of factors. These include the limits imposed by the response scale, issues of reverse causality and the possible impact of "culture" on trust estimates. These factors are briefly described below, together with an evaluation of their severity.

Box 5.1. **Demand effects and question framing**

Apart from differences between groups or time points, trust items may be subject to another type of difference. Namely, stakeholders might also be interested in assessing whether interpersonal trust depends on the type of group to be trusted (people in general, people you know personally, friends, etc.) and whether different public institutions (the national government, public agencies providing different services, the judiciary, state and local governments) are trusted to varying degrees. However, one should be careful not to over-interpret the size of differences between targets of trust when respondents are asked to provide their ratings for a long list of differently framed questions. Indeed, demand effects (changes in behaviour by respondents due to cues about what constitutes appropriate behaviour by the survey or experimental design) might lead to the assumption that slightly different information is wanted for each item. In other words, being asked repeatedly about similar – yet related constructs – might lead respondents to adjust their answers, thus inflating the magnitude of differences between trust in a range of social groups or institutions. The survey design should always be reviewed to assess whether it could have encouraged demand effects, including the order in which different types of people and institutions are listed, and whether any adjustments are needed. For instance, as outlined in the prototype module instructions in Annex 2, in cases where assessing trust in different institutions is important, the order in which questions are presented should be randomised across respondents.

One practical challenge that can influence the size of differences for trust items is the design of response scales. The specific features of the response scale influence which responses are theoretically possible and how responses are distributed. Since trust data are

collected with bounded scales with a limited number of response categories, the average trust measure can never exceed the top response category (unlike some measures where the scale is unbounded, such as income). Nevertheless, despite this theoretical limit, most countries are currently far from the top of the scale: for example, while, in the 2013 EU-SILC data, Denmark has the highest score (8.3 out of 10) for interpersonal trust, the overall sample mean lies at 5.8. In the case of institutional trust, the top scores in the EU-SILC 2013 data do not exceed 8 points. Looking at global trust measures, in some countries (e.g. Singapore, Kazakhstan, Vietnam) between 81% and 91% of respondents in the Gallup World Poll state that they trust government based on a dichotomous question item – however, these findings may reflect the small sample size and the influence of political considerations on respondents (this underscores the importance of managing sensitivity concerns through sound survey design and mode, as described in Chapter 3). Overall, apart from a few outliers, there seems to be quite a lot of "upwards space" left until countries hit the maximum ceiling allowed by the respective response scales.

The possibility of two-way causality between trust and its determinants may also influence and limit the size of the differences likely to be observed. Two-way causality occurs where there are reciprocal and causal relationships between two variables, running in both directions. For example, some researchers have argued that membership in voluntary associations increases interpersonal trust, while others have asserted that one's level of trust in others also drives the likelihood of volunteering (Stolle, 1998; Uslaner, 2000). Two-way causality points to the importance of selecting an appropriate time frame when examining changes in trust, based on what is known about the variables in question and the possible causal pathways through which they might take effect. Although longer time-frames might be required to detect significant changes in interpersonal trust data, measures of trust in institutions in particular might be relatively bumpy over short time periods. For example, Chanley (2002) found that trust in the US federal government experienced a sharp rise after 11 September 2001, a phenomenon Deaton (2012) refers to as short-term "cognitive bubbles" in reaction to significant national events. Further research will be needed to identify which events have a short-term impact and which have a longer-term impact on trust levels. For example, levels of institutional trust in Europe, particularly in politicians and the government, have so far not bumped back to pre-financial crisis levels. In general, time-series data on trust should be examined over longer time periods to identify meaningful changes associated with persistent societal and economic shifts.

Cultural differences in how respondents understand, process and reply to subjective questions, and the frames of reference they rely on when doing so, may also drive difference in international comparisons of trust. Especially in the case of interpersonal trust, it may seem puzzling that countries at similar levels of economic development report quite different mean levels of trust. Figure 5.4 above portrayed the mean distribution of interpersonal trust across the EU-24 countries: average country scores range all the way from 5 (France) at the bottom to 8.3 (Denmark) at the top end of the distribution. Countries with relatively high levels of GDP, notably Germany and France, are located near the low end of the distribution. In fact, the low ranking of France (within Europe) holds not only for trust in other people, but also for other self-reported measures such as subjective well-being and trust in the market. Senik (2011) attributes this "French unhappiness puzzle" to unique cultural factors and mental attitudes of French people.

Chapter 3 already described the evidence around cultural response styles and the methodological steps in survey design that can be taken to reduce the risk that scales and

questions might be understood differently by respondents. This section focuses on how to interpret and deal with cultural biases, once the data have been collected. In the following, possible sources of cross-country differences in average trust levels are highlighted and methods to "correct" data for cultural bias are briefly introduced – while also raising the question of whether such corrections should be conducted at all. Before attributing differences in average trust between countries at similar levels of economic development to "cultural bias", it is important to remember that these differences may also have multiple sources. A useful distinction can be made between "cultural impact", i.e. genuine sources of variance between cultures, and "cultural bias", i.e. inter-cultural differences stemming from measurement artefacts (Van de Vijver and Poortinga, 1997; Exton et al, 2015). The potential sources of cross-country variance described below all carry different implications for the validity of between-country comparisons of trust data, as well as for any actions that one might take to mitigate the impact of cultural bias.

Sources of cultural impact

On the one hand, there are country-specific differences for which one would not necessarily want to "correct". For example, one source of international variance in trust scores, in addition to economic variables, involves cross-country differences in the *social and political context* and other life circumstances of residents, all of which might impact upon trust. These drivers may or may not be related to culture: they include social relationships, the homogeneity of society in terms of income and ethnic diversity, religion, unemployment, the rule of law, crime, corruption, and the type and quality of institutional arrangements (Alesina and Ferrara, 2000; Algan and Cahuc, 2013; Guiso et al., 2006; Stolle et al., 2008; Rothstein and Uslaner, 2005; Jordahl, 2007).

The socio-demographic structure of the particular sample in each country may also contribute to differences in reported trust: trust is shaped by individual background characteristics, such as a respondent's age, gender, income, education and employment status, and features of the community where he or she lives. For example, there is a robust relationship between both interpersonal and institutional trust and educational status (Stolle et al., 2008; Helliwell and Wang, 2010; Carl and Billari, 2014). Theoretical reasons that have been put forward for this include the fact that education – and the higher level of income associated with it – empowers people, enabling them to make their own choices and to accomplish their goals and expectations (Hudson, 2006). An advanced education is also likely to be linked with a better understanding of how public institutions function.[7] Moreover, people who are better off financially and more educated are likely to enjoy more opportunities and channels to take part in society (e.g. through volunteering and political participation), which is conducive to developing and maintaining larger and more diverse social networks (Helliwell and Putnam, 2007; OECD, 2015).[8] Unemployment is another socio-economic factor that has been associated with the degree of an individual's trust both in other people and in institutions, with the unemployed exhibiting lower trust compared to their fellow citizens (Chabanet, 2007; Bârgăoanu et al., 2015). It has also been found that interpersonal trust increases slightly with age (Stolle et al., 2008; Putnam, 2000; Tokuda et al., 2008; Li and Fung, 2012; Clark and Eisenstein, 2013). While the exact reasons why older people report higher trust in others are under debate, one potential explanation for age-related increases in trust is that older adults are more motivated to give back to others, therefore believing them to be good and trustworthy in return (Poulin and Haase, 2015). Generally, it is very important to examine each sample, including across countries, to understand how its composition might have driven effect sizes.

There may also be differences between countries in how people *feel* about trusting other people or institutions. These differences in feelings are influenced by many features, such as an individual's reference group, past experiences, and the historical roots of the country both of origin and of current residence (which can set collective *frames of reference*). "Frame-of-reference effects" refer to differences in the way respondents answer survey questions based on their own life experiences as well as on their knowledge about the experiences of others, both within and outside their comparison group (Sen, 2002; Ubel et al., 2005). These experiences set the frame of reference, relative to which a respondent's own current circumstances and feelings are evaluated. The frame of reference may contribute to appraisal styles that influence the connection between objective life circumstances and subjective feelings – for example, the degree of optimism or pessimism that individuals feel about the actions of other people. Frames of reference produce real differences in levels of trust, rather than simply differences in how people report those feelings. They thus do not bring into question the validity of trust measures per se. Evidence from other self-reported measures, such as subjective well-being, suggests that while framing effects may influence the size of differences between groups and countries, they are not sufficiently large to prevent the detection of the impact of life circumstances (OECD, 2013). This source of cross-country variance can reflect cultural impact rather than bias and might even add to the predictive validity of the overall trust measure (i.e. in its association with real-life trusting behaviour towards other people and institutions).

Sources of cultural bias

On the other hand, *linguistic differences* and *cultural response styles* are likely to add cultural bias to the data, reducing their overall validity and predictive ability. Linguistic differences are likely to play a key role when trust constructs are not perfectly translatable across languages – for instance, many languages do not make the conceptual distinction in English between *trust* and *confidence*. The translatability of constructs is also likely to be more challenging when response scales are verbal rather than numerical (Veenhoven, 2008). Cultural response styles, described in Chapter 3, refer to group differences in scale use or differences in how individuals *report* their feelings. For example, a "modesty" or moderate-responding bias might have a downward influence on self-reports, without having a negative impact on private feelings of trust. Similarly, tendencies towards *extreme responding* (i.e. using scale end points) or more socially desirable responding could imply differences in modes of cultural expression, rather than substantive differences in the trust actually experienced. Both linguistic differences and cultural response styles represent sources of bias that should be minimised at source through survey design (Chapter 3) or translation (Chapter 4) or by adjusting the data *ex post* to correct for the bias.

Methods for examining and removing cultural bias include the use of objective outcomes as counterfactuals in the analysis (here, experimental measures of trust might be useful in the future), fixed-effects models to control for country and regional characteristics, or vignettes to measure the different ways in which individuals and/or cultures may interpret or benchmark the same survey question (OECD, 2013). Another technique for investigating the effects of culture that is gaining prominence among scholars is the use of migrant data, i.e. the comparison of response styles of natives and migrant respondents within the same country for particular outcomes of interest (Senik, 2011; Exton et al., 2015).

However, none of these approaches has yet convincingly distinguished between a substantive cultural impact and cultural bias. The relatively small number of countries

sampled in the existing research also makes it difficult to extrapolate results more widely – meaning that there is little that can be said even about the expected magnitude of cultural effects, particularly at a global level. Access to further high-quality data on trust from large and nationally representative samples will help to shed light on the issue of what proportion of average-level cross-country differences can be attributed to cultural biases. This information will help to determine whether the benefits of data adjustments outweigh the risk of removing the influence that all unmeasured country differences (including the influence of a country's policy environment, social networks and a wide range of valid cultural differences) have on how trust assessments are formed and maintained. A further practical limitation in using vignettes and migrant data to correct country averages of trust is that the impact of culture in the data cannot be quantified in absolute terms – rather, it is always *relative* to other countries in the sample. This provides a further challenge if the goal is to adjust national-level data to provide *culture-free* estimates; it implies that only a large and representative global sample could be used as a basis for such adjustments. Given these limitations and the current state of evidence, these Guidelines recommend against using methods to correct national trust data for cultural influences for the time being.

5.4. Analysing trust data

Moving beyond descriptive presentations of trust metrics is of interest to both the general public and policy makers for several reasons. First, interpersonal trust and trust in institutions are valuable outcomes in their own right. Especially when identifying vulnerable groups and international benchmarking are core elements of monitoring trust, a better understanding of what causes, maintains or destroys trust can help explain some of the observed differences between countries and groups. Analysing the drivers of trust can then support both the identification of those areas where appropriate policies could raise trust as well as the informed appraisal of various policies that might have unintended impacts on trust.

Second, trust also impacts on a multitude of other outcomes of interest, including economic growth, subjective well-being, health status, crime levels and the willingness to participate in the democratic process. This makes it imperative to better understand how to improve trust and to unpack the exact pathways of how trust drives people's well-being.

Different methods of analysis are described below, before looking at selected examples of how both interpersonal and institutional trust has been analysed in practice. The section closes with some words of caution about the analytical challenges that might arise when working with trust data.

Data requirements and study design

Better understanding the drivers and impacts of trust involves a process of identifying the variables that have causal relationships with trust or on which trust has a causal effect and examining the mechanisms through which these effects take place.

Before presenting a range of basic analytical methods applicable to trust data, a few comments on general data requirements are warranted. Preferably, surveys of trust are already designed with some idea of the final analysis in mind and include a wide range of covariates to draw upon, including a number of standard demographic and control variables and measures of potential drivers of trust. Chapter 4 elaborates further on which data to collect alongside trust during the measurement stage. Since any analysis of drivers requires access to micro-level data, the trust datasets gathered by NSOs should ideally be

anonymised and made publicly accessible to government analysts, researchers and organisations that have an interest in informing policy and public debate.

The nature of the study design with which the data was acquired is a key determinant of the degree to which researchers are able to make inferences about the causality of relationships between variables. The "gold standard" for establishing causality in social research is the so-called experimental design, or randomised control trials (RCTs). RCTs, well known for their use in trials of new medications, involve the random allocation of individuals to control groups (not receiving an intervention) and treatment groups (receiving intervention A, B, C, etc.). The impact of the intervention is then established by comparing the outcomes of the treatment group with those of the control group, effectively presenting a "counterfactual", or "what would have happened to the treatment group in the absence of the intervention". The use of RCTs in policy research, especially with regard to trust-relevant interventions, is rare and expensive. Furthermore, they are often open to ethical criticism (whenever it can be difficult to justify why treatment should be withheld from one group); also, it is not always clear whether the findings established in a small-sample RCT will hold when an intervention is scaled up or implemented in another or broader context.

A second-best option is quasi-experimental study design. One form of a quasi-experimental pre-test/post-test design is regression discontinuity design, in which the causal effects of interventions are elicited by setting a cut-off or threshold above or below which an intervention is assigned. Regression discontinuity designs are appropriate when randomisation is not feasible, but still require that the intervention itself is under the control of the researchers. Another form of a quasi-experiment is the "natural experiment", in which individuals exposed to the experimental and control conditions are determined by nature or by other factors outside the control of the researchers, but where the process governing the exposure arguably resembles random assignment. However, natural experiments are difficult to find (as they tend to happen by chance), and many times not all baseline data of interest were collected or are available via administrative sources. In a way, international comparisons between countries, in which a particular intervention was implemented in one country but not in others, are a particular case of a natural experiment. However, it is generally very difficult to infer causality from international comparisons, given the variety of uncontrolled differences between countries in terms of both sample characteristics and other variables of interest. This is especially the case when relying on cross-sectional (rather than longitudinal) data. Ideally, researchers would work with panel data (longitudinal surveys collecting repeated measures for the same person over time), as this offers an opportunity to explore whether a change in a given determinant is associated with a subsequent change in a person's reported trust. Data quality, and the generalisability of findings, are further enhanced when panel data stem from large and representative samples such as those obtained by NSOs.

However, since large-scale comparable panel data are rarely available, most studies concerned with the impacts and drivers of trust must rely on cross-sectional datasets. Strictly speaking, such analyses are concerned with identifying covariates rather than causality. Nevertheless, even when direct causal inferences cannot be made with cross-sectional data, evidence from other sources about the direction of causality can be used to enrich the interpretation of the results.

Methods of analysis

Apart from the nature of the research question, the most appropriate method for the analysis of trust depends largely on the type of data collected and the method of collection.

The simplest test for the strength of a relationship or association between two variables is a bivariate correlation. The Pearson or product-moment coefficient can be calculated when the data are assumed to be normally distributed and the expected relationship between them is linear; Spearman's Rank and other non-parametric tests are available for ordinal data and non-linear relationships. Partial correlation enables examining the relationship between two variables while removing the effect of other variables. Correlations indicate the possible existence of a predictive relationship between two variables, but they do not imply causation.

For a more thorough examination of the impacts and drivers of trust in cross-sectional, international and longitudinal studies, regression analysis is widely used. Regression is a correlation-based statistical technique that examines how well a set of explanatory or independent variables can predict a given dependent variable, i.e. the chosen trust measure. Regression allows assessing the impact of several independent variables simultaneously in one model, and can be used even when explanatory variables are correlated with one another. However, the "best" regression solution (in terms of variance explained per independent variable) is produced when each independent variable is strongly correlated with the outcome variable but uncorrelated with other variables, whether these other variables are included or excluded from the model. Where curvilinear relationships are expected, as in the case of the U-shaped relationship between age and trust, squared values are typically used in regression models. For income, which is expected to have an asymptotic relationship with many outcome variables, values are often transformed into logs.

A range of regression models are available, depending on the nature of the independent variable: linear regression models are recommended for continuous variables, while ordinal and dichotomous outcomes are usually analysed with Probit or Logit models. Since trust measures are not continuous, a Probit or Logit model is the most appropriate in studies where trust is the outcome of interest. However, the output of linear regression models is generally easier to interpret, and in many cases there are few differences between estimates based on linear regression and Probit/Logit models (Diener and Tov, 2012). Therefore, it is recommended that both types of analysis are run when dealing with trust data, and that linear regression estimates are reported when the results do not differ.

More advanced methods of analysis, not described in detail here, include multilevel structural equation modelling, which allows for combining micro-level and macro-level information, and propensity score matching, which estimates the effect of an intervention by accounting for the covariates that predict receiving the treatment.

How to assess whether the results are significant

Associations between variables can be analysed by examining their correlation coefficients (denoted as r). These range from -1 to +1, with -1 implying a perfect negative linear association and +1 a perfect positive linear association. The square of the coefficient (or r^2) denotes the share of the variation in one variable that is related to the variation in the other. Thus, an r^2 of 0.36 (i.e. r of 0.60) means that 36% of the variance in the dependent variable is explained by the variance in the independent variable. The statistical significance of a correlation coefficient indicates the likelihood that the coefficient would be found in a sample by chance when no significant association actually exists between the variables.

In regression-based analyses, the overall model "fit" of the observed data is described in terms of the proportion of variance in the dependent variable that is explained by the variance in the independent variables (the overall multiple-correlation coefficient, or R^2).

Statistical significance is used to indicate whether the overall model provides a better-than-chance prediction of the dependent variable. In order to further understand how each independent variable contributes to the prediction of the dependent variable, one examines the set of regression coefficients for the independent variables. In linear regression, the size (and sign) of the coefficient for each independent variable indicates how much the dependent variable is expected to increase (if positive) or decrease (if negative) when the independent variable increases by one unit, while holding all the other independent variables constant.

The analysis of trust data in practice

The results from selected econometric analyses that have been carried out with trust data are described below to illustrate the different ways in which trust data have been modelled as well as the effect sizes that have typically been produced so far. The description distinguishes between studies that have considered trust as an outcome variable (i.e. where the analysis is concerned with establishing the determinants of trust) and those that have viewed trust as an input variable (i.e. where the analysis aims at quantifying the impact of trust on other outcomes of interest). In this context, the use of variables at different units of analysis, namely at the individual or community/country-level, is highlighted as well.

Trust as an outcome

Anderson and Tverdova (2003) provide an example of a study that is concerned with trust in institutions (in their case, trust in civil servants, expressed on a 1-5 scale) as an outcome. The authors combine individual-level trust data from the International Social Survey Program (ISSP) with information on respondents' political allegiance (the ISSP survey also includes an item about which party the individual voted for in the last election) as well as with Transparency International's country-level Corruption Perception Index (CPI) for 16 countries at different maturity stages of democracy. Controlling for various system and individual-level factors, including current macroeconomic performance, economic development, level of democracy, political interest, electoral participation, socio-economic status and standard demographic variables, the authors use multilevel structural equation-modelling to establish the impact of corruption and political allegiance on trust. Their findings indicate that individuals in countries with higher levels of corruption have less trust in civil servants, and that respondents who support the majority political party are significantly more trusting of civil servants than those in the minority. For example, a typical respondent in a country where corruption is absent scores 4.26 on the 5-point scale measuring respondents' trust in civil servants. In contrast, a respondent in a country in the most corrupt category scores 2.76, while those in a country in the mid-range of the corruption scale (=5) rate civil servants 3.33. Looking at the effect of political allegiance in the majority or minority party on trust in civil servants, those supporting the majority party score 3.98 in the average country, while those favouring the minority party score 3.83.

An example of a study that looks at the determinants of interpersonal trust is the research by Alesina and La Ferrara (2000) on US communities. Using micro-level trust data from the US General Social Survey (GSS) from 1974-94 as dependent variable, the authors rely on multiple Probit regression analysis (since the trust item is measured on a binary scale) to test for the impact of a rich set of independent variables: these include individual characteristics and experiences also featured in the GSS (education, income, marital status,

age, gender, religious affiliation, traumatic experiences such as divorce, disease, accidents and financial misfortune) as well as community characteristics drawn from administrative data (income inequality, racial and ethnic homogeneity, crime level). The authors find that the factors reducing trust most significantly are: a recent history of traumatic experiences; belonging to a group that historically feels discriminated against, such as minorities (Blacks in particular); and, to a lesser extent, being a woman, being economically unsuccessful in terms of income and education, or living in a racially mixed community and/or in one with a high degree of income disparity.[9] For example, in econometric terms, an increase in the Gini index by one standard deviation decreases the likelihood of trust by 2.5 percentage points. Interestingly, when the authors repeat their analysis with trust in a range of public institutions as dependent variable, they find no effect of community racial and income heterogeneity. These findings suggest that these factors influence interpersonal interactions but not trust in institutions.

Trust as a driver of other outcomes

Interpersonal trust has been analysed intensively not only as an outcome, but also as a determinant of other variables of interest. In particular, interpersonal trust has been linked in cross-country studies to income per capita and economic growth (Putnam et al., 1993; Knaack and Keefer, 1997; Ahn and Hemmings, 2000; Temple, 2000).

For example, Algan and Cahuc (2013), using a sample of 106 countries[10] over the years 1981 to 2008, regress the log of income per capita on average general social trust, including controls for education, ethnic fractionalisation and population size. Their results indicate that increasing trust by one standard deviation leads to a rise in income per capita of 0.18, or 2% of the sample mean. Even when introducing additional control for institutional history and quality, interpersonal trust remains significant at the 5% level. The authors replicate these findings using data from 800 regions around the world, relying on weighted regressions using the number of individuals polled per region. In this case, however, the positive and significant correlation between trust and GDP per capita disappears once country fixed effects are introduced into the model, suggesting that cross-country heterogeneity rather than within-country variation in interpersonal trust and income per capita drives the result. Apart from income per capita, the authors also examine the relationship between trust and economic growth. Regressing average annual growth between 1990 and 2009 on average trust between 1981 and 1990 for a 52-country sample as featured in the World Values Survey, and controlling for the country's investment level, their findings yield a correlation coefficient of 0.48 between trust and growth, significant at the 5% level. The overall R^2 of the model of Yann and Cahuc is 0.706, i.e. explaining 70% of the variance in economic growth.

Apart from economic outcomes, interpersonal trust has also been found to predict a range of social outcomes, such as different dimensions of health status and health-related behaviour (Lochner et al., 2003; Lindström, 2005; Brown et al., 2006; Poortinga, 2006; Petrou and Kupek, 2008), crime rates (Buonanno et al., 2009) and subjective well-being (Helliwell and Wang, 2010). Boarini et al. (2012) explore the determinants of subjective well-being across OECD countries and conclude that country-average trust predicts individual life satisfaction at the 5% significance level. Drawing on Gallup World Poll data, the authors' linear regression, controlling for demographics, socio-economic status and other variables related to well-being, estimates that the impact on life satisfaction of a one-unit change in aggregate average interpersonal trust is equivalent to multiplying current income by 1.23.

Generally, studies examining either the causes or the impacts of trust illustrate that the choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – can make an important difference to the results. Most often, it is community and country characteristics that influence trust levels, while aggregate trust levels have been shown to affect individual-level outcomes, such as life satisfaction and health status. This is, in some way, good news for policy makers, since it suggests that policy interventions at the aggregate level (local, regional, national) are the most relevant.

Challenges to interpreting coefficients

A few words of caution about some analytical challenges in econometric exercises are warranted. While these challenges are not unique to studies examining the drivers or impacts of trust, they should be kept in mind when interpreting results, particularly if these are used to inform policy.

First, normally the size of regression estimates can be interpreted as the effect on the dependent variable of a one-unit change in the independent variable. However, regression coefficients can be affected if the independent variables in the equation are strongly correlated amongst themselves. The literature distinguishes between mediation, confounding and suppression effects resulting from the inclusion of additional independent variables (Dolan et al., 2007). Furthermore, if several measures of the same driver are included in the model, their intra-correlations can crowd one another out, to the limit case where an otherwise relevant driver can fail to reach significance and hence be overlooked (Boarini et al., 2012 refer to this scenario as "over-identification"). Therefore, the decision of which and how many variables to include in the analysis should always be informed by a clear theoretical structure and by an understanding of the hypothetical causal pathways between the different factors.

Second, estimates will be affected by endogeneity problems when the variable of interest is correlated with the model error term, an issue that is often referred to as the "omitted variable problem". In this case, variables that are omitted, i.e. not included in the model, but which are causally related to both outcome and predictor variables, can make it seem as if there is a significant statistical relationship between the latter two, although none exists in reality. In a hypothetical example, a relationship between interpersonal trust and community-level ethnic diversity might be caused by community-level income inequality, which it will be important to include as a measure in the model as well.

Third, the ability to understand the direction of causality between trust and other variables is strongest when experimental, quasi-experimental or longitudinal panel data are available. More often than not, this will not be the case, and analyses have to rely on cross-sectional data. In this case, the results need to be interpreted alongside evidence about the causal direction from other sources. One method often used to overcome issues of reverse causality in regression analyses is to include instrumental variables. An instrumental variable is one that has a direct association with the independent variable in question (e.g. trust), but not with the outcome of interest (e.g. GDP growth). Although it can be difficult to identify appropriate instrumental variables, several researchers have exploited the inherited trust of US immigrants as an instrument for trust in their origin country in cross-country regressions of GDP growth (Sangier, 2010; Algan and Cahuc, 2010).

Very importantly, the possibility of shared method variance should always be kept in mind when interpreting the results of analysis of trust data. Shared method variance refers

to variance that is attributed to the measurement method, rather than the constructs of interest. In the case of trust, the main concern is that, when drivers are also measured through self-reported data, self-report biases (including social desirability biases, response styles, cultural bias, etc.) can inflate the estimated impact of self-reported drivers relative to those measured through other means (e.g. objective observations). Questions that have very similar response formats (e.g. 0-10 scales) are especially likely to have correlated errors.

When comparing the effects and impacts of trust, particularly in cross-sectional data, it is therefore important to consider how each of those variables was measured. Whenever possible, the use of longitudinal data is encouraged, as individual fixed effects can be controlled for in these analysis formats. Alternatively, non-self-reported measures of constructs of interest should substitute self-reported items if data are available – for example, behavioural measures of trust collected through behavioural games.

5.5. Conclusion

This chapter has addressed different ways of working with trust data *after* it has been collected, ranging from reporting interpretation all the way to analysis. It has outlined best practice where available, raised awareness of challenges to interpretation and analysis, where relevant, and pointed to gaps in current knowledge where they exist. In doing so, the chapter aims to serve as a useful reference handbook for data analysts, researchers and journalists interested in outputting, reporting on and analysing trust data. The key messages of the chapter can be summarised as follows:

- Analysts tasked with reporting trust data have an important communications role to play and should take into account the intended target audience and its needs. The front page of statistical releases should provide a quick overview for the general public and policy makers and focus on reporting levels of a single headline measure alongside a brief media commentary. In contrast, the second and third pages of statistical releases should cover trust data in greater detail (e.g. trends in and distributions of trust, different types of trust) for those who want to dig deeper. The third page of data releases is designed for researchers and policy analysts who want to carry out analyses themselves and require access to the trust microdata and the survey instruments.
- There are several ways to output trust data, each with unique pros and cons. Trust levels can be presented by reporting frequencies in each category, the proportions above or below thresholds, or central tendency measures (i.e. mean, median, mode). Some rules of thumb for best reporting practice include refraining from arbitrary labels for thresholds (e.g. *high*, *low*) and complementing mean levels with information about the distribution of data, such as the standard deviation. Changes over time can be monitored by tracking changes in mean trust through time series or by calculating changes in the mean score over various points in time. Group differences can be examined by presenting group differences over time, relative to a given threshold, or by showing the (absolute or percentage) differences in the proportion of respondents who have selected a specific answer. Both sample size and standard errors should be reported alongside group means.
- Essential questions for the interpretation of trust data deal with what should be considered a *small* or a *big* difference between observations in real-life terms, and the extent to which observed differences are influenced by measurement artefacts and errors. While the chapter provides an initial attempt to document the magnitude of differences (between population subgroups, between countries, over time) encountered up to now

with existing data, many gaps remain, and knowledge on the overall data universe of trust will remain limited until higher quality and more frequent data become available.

- The magnitude of differences between and within observations (over time), can be influenced by a variety of factors that should be factored into any interpretation exercise. These include the limits imposed by the response scale, issues of reverse causality and the possible impact of culture on trust estimates. This chapter stressed the point that, since it is very difficult in practice to distinguish between cultural impact and cultural bias, and the methods available to remove cultural bias face significant limitations, it is generally not recommended to correct trust data for cultural influences.
- Sound analysis of trust data requires access to data from which causal inferences can be made and that the relevant covariates, including standard demographic and control variables, are ideally collected in the same survey. The chapter has presented a range of analytical methods applicable to trust data, providing examples of studies that have relied upon trust as either an input or output variable. These studies illustrate that the choice of the unit of analysis – i.e. whether to consider variables at the individual or country level – matters, that community characteristics often influence trust levels, and that aggregate trust levels influence individual-level well-being outcomes.
- Lastly, the chapter has highlighted the importance of keeping common econometric challenges in mind when working with trust data. These include omitted variable bias, over-identification, reverse causality and shared method variance.

Notes

1. The mean, median, and mode are the most common measures of central tendency. The mean refers to the sum of all measurements divided by the number of observations in the dataset; the median is the middle value that separates the upper 50% from the bottom 50% of the dataset; and the mode is the most frequent value in the dataset. The median and the mode are the only measures of central tendency that can be used for ordinal data, whereas the mode is the only central tendency measure applicable to categorical data.
2. Ordinal data are those measured on scales where the intervals between scale points are not assumed to be equal, but there is an underlying sequence or rank order. For example, we assume that a 5 is lower than a 6 and a 6 is lower than a 7, but we do not assume that the distance between 5 and 6 is equivalent to the distance between 6 and 7. Linear regression relies on continuous variables, where cardinality is assumed, i.e. the size of the number on a scale is expected to have a direct linear relationship with the amount of the variable in question. Tabachnick and Fidell (2001), however, note that, in the social sciences, it is common practice to treat ordinal variables as continuous, particularly where the number of categories is large – e.g. seven or more – and the data meet other assumptions of the analysis.
3. Note by Turkey: The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".
4. Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.
5. See Notes 3 and 4.
6. The importance of considering changes in individuals' levels of trust is highlighted by Bilson et al. (2017). The authors, using panel data from the Household, Income and Labour Dynamics in Australia (HILDA) survey from 2005-2014, investigate the effect of individual income on interpersonal trust, demonstrating the importance of accounting for individual-level fixed effects: the income coefficient switches from positive and statistically significant, in a pooled regression, to negative and

statistically significant, in a fixed-effects panel model. These findings underscore the need for individual-level panel data for all trust measures.

7. However, some researchers have also found that the education gradient in trust, especially in the US context, has generally diminished over time (see Dalton, 2005).
8. It should be noted that there is some debate about the direction of causality between educational attainment and trust, with some authors asserting that education impacts trust, and others claiming that trust causes differences in education (see Bjørnskov, 2006, for a discussion).
9. Putnam (2007) reports similar effects of community contextual variables on individual levels of interpersonal trust in the United States, including significant negative impacts of poverty rates, non-violent crimes, an index of ethnic homogeneity, and population density at the census tract level. Bjørnskov (2006), for an international sample, and Helliwell and Wang (2010), focusing on Canada, also report similar findings. Nevertheless, these results should not be generalised lightly: they need to be interpreted with the historical context of each country in mind, as well as while recognising that no consensus currently exists on how to define and measure ethnicity and ethnic fragmentation.
10. The authors use binary measures of trust drawn from the World Values Survey (1981-2008), the European Values Survey (1981-2008) and the Afrobarometer (2005).

References

Ahn, S. and P. Hemmings (2000), "Policy influences on economic growth in OECD countries: An evaluation of the evidence", *OECD Economics Department Working Papers*, No. 246, OECD Publishing, Paris, *http://dx.doi.org/10.1787/581718238837*.

Alesina, A. and E.L. Ferrara (2000), "The determinants of trust", *National Bureau of Economic Research*, No. w762, *www.nber.org/papers/w7621*.

Algan, Y. and P. Cahuc (2013), "Trust, growth and well-being: New evidence and policy implications", *IZA Discussion Paper,* No. 7464.

Algan, Y. and P. Cahuc (2010), "Inherited trust and growth", *The American Economic Review,* Vol. 100(5), pp. 2060-2092.

Anderson, C.J. and Y.V. Tverdova (2003), "Corruption, political allegiances, and attitudes toward government in contemporary democracies", *American Journal of Political Science*, Vol. 47, pp. 91-109.

Bârgăoanu, A., L. Radu and D. Varela (2015), *United by or Against Euroscepticism? An Assessment of Public Attitudes towards Europe in the Context of the Crisis*, Cambridge Scholars Publishing, Cambridge.

Bilson, J.R., M. Jetter and I. Kristoffersen (2017), "Gender differences in the link between income and trust levels: Evidence from longitudinal data", *IZA DP,* No. 10585.

Bjørnskov, C. (2006), "Determinants of generalized trust: A cross-country comparison", *Public Choice*, Vol. 130, pp. 1-21.

Blanton, H. and J. Jaccard (2006), "Arbitrary metrics in psychology", *American Psychologist*, Vol. 61, pp. 27-41.

Boarini, R. et al. (2012), "What makes for a better life? The determinants of subjective well-being in OECD countries – Evidence from the Gallup World Poll", *OECD Statistics Working Papers*, No. 2012/03, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5k9b9ltjm937-en*.

Brown, T.T. et al. (2006), "The empirical relationship between community social capital and the demand for cigarettes", *Health Economics*, Vol. 15 (11), pp. 1159-1172.

Buonanno, P., D. Montolio and P. Vanin (2009), "Does social capital reduce crime?", *Journal of Law & Economics*, University of Chicago Press, Vol. 52(1), pp. 145-170.

Carl, N. and F.C. Billari (2014), "Generalized trust and intelligence in the United States", *PLoS ONE*, Vol. 9(3).

Chabanet, D. (2007), "Chômage et exclusion sociale: L'échec européen", *Politique Européenne*, Vol. 2, pp. 157-187.

Chanley, V.A. (2002), "Trust in government in the aftermath of 9/11: Determinants and consequences", *Political Psychology*, Vol. 23, pp. 469-483.

Clark, A.K. and M.A. Eisenstein (2013), "Interpersonal trust: An age-period-cohort analysis revisited", *Social Science Research*, Vol. 42, pp. 361-375.

Dalton, R.J. (2005), "The social transformation of trust in government", *International Review of Sociology*, Vol. 15(1), pp. 133-154.

Deaton, A. (2012), "The financial crisis and the well-being of Americans; 2011 OEP Hicks Lecture", *Oxford Economic Papers*, No. 64, pp. 1-26.

Diener, E. and W. Tov (2012), "National accounts of well-being", *Handbook of Social Indicators and Quality of Life Research*, pp. 137-157, K.C. Land, A.C. Michalos and M.J. Sirgy (eds.), Dordrecht: Springer.

Dolan, P., T. Peasgood and M. White (2007), "Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being", *Journal of Economic Psychology*, Vol. 29, pp. 94-122.

Eurostat (2015), *European Union Statistics on Income and Living Conditions (EU SILC)*, *http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en* (accessed 23 March 2017).

Exton, C., C. Smith and D. Vandendriessche (2015), "Comparing happiness across the world: Does culture matter?", *OECD Statistics Working Papers*, No. 2015/04, OECD Publishing, Paris, *http://dx.doi.org/10.1787/5jrqppzd9bs2-en*.

Gallup (2014), *Gallup World Poll, www.gallup.com/services/170945/world-poll.aspx* (accessed 06 February 2017).

Guiso, L., P. Sapienza and L. Zingales (2006), "Does culture affect economic outcomes?", *The Journal of Economic Perspectives*, Vol. 20(2), pp. 23-48.

Helliwell, J.F. and R.D. Putnam (2007), "Education and social capital", *Eastern Economic Journal*, Vol. 63(1), pp. 1-19.

Helliwell, J.F. and S. Wang (2010), "Trust and well-being", *National Bureau of Economic Research*, Working Paper No. 15911.

Hudson, J. (2006), "Institutional trust and subjective well-being across the EU", *Kyklos International Review for Social Sciences,* Vol. 59(1), pp. 43-62.

Jordahl, H. (2007), "Inequality and trust" *Working Paper Series, No.* 715, Research Institute of Industrial Economics.

Knack, S. and P. Keefer (1997), "Does social capital have an economic payoff? A cross-country investigation", *Quarterly Journal of Economics*, Vol. 112 (4), pp. 1252-88

Li, T. and H.H. Fung (2012), "Age differences in trust: an investigation across 38 countries", *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*.

Lindström, M. (2005), "Social capital, the miniaturization of community and high alcohol consumption: A population-based study", *Alcohol and Alcoholism*, Vol. 40 (6), pp. 556-562.

Lochner, K.A. et al. (2003), "Social capital and neighborhood mortality rates in Chicago", *Social Science & Medicine*, Vol. 56, pp. 1797-1805.

New Economics Foundation (2009), contributing authors: J. Michaelson, S. Abdallah, N. Steuer, S. Thompson and N. Marks, "National accounts of well-being: Bringing real wealth onto the balance sheet", *New Economics Foundation*, January, London, *www.nationalaccountsofwellbeing.org/learn/download-report.html*.

New Zealand State Services Commission (2014), *Kiwis Count Survey* (database), *www.ssc.govt.nz/kiwis-count-datasets* (accessed 20 April 2017).

OECD (2015), Education at a Glance 2015: OECD Indicators, OECD Publishing, Paris, *http://dx.doi.org/10.1787/eag-2015-en*.

OECD (2013), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264191655-en*.

ONS (2016), "Statistics on trust for methodological testing from the opinion's survey, October 2015 to May 2016", released 10 November, Office for National Statistics UK, ONS, Newport, *www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/adhocs/006326statisticsontrustformethodologicaltestingfromtheopinionssurveyoct2015tomay2016*.

Pew Research Center (2016), *Beyond Distrust: How Americans View their Government*, *www.people-press.org/2015/11/23/1-trust-in-government-1958-2015/* (accessed 18 February 2017).

Petrou, S. and E. Kupek (2008), "Social capital and its relationship with measures of health status: Evidence from the Health Survey for England 2003", *Health Econ.*, Vol. 17, pp. 127-143.

Poortinga, W. (2006), "Social relations or social capital? Individual and community health effects of bonding social capital", *Social Science & Medicine*, Vol. 63(1), pp. 255-270.

Poulin, M.J. and C.M. Haase (2015), "Growing to trust", *Social Psychological and Personality Science,* Vol. 6(6), pp. 614-621.

Putnam, R.D. (2007), "E Pluribus Unum: Diversity and community in the Twenty-First Century – the Johan Skytte Prize Lecture", *Scandinavian Political Studies*, Vol. 30, pp. 137-74.

Putnam, R.D. (2000), *Bowling Alone: The Collapse and Revival of American Community*, Simon & Schuster (eds.), New York.

Putnam, R.D., R. Leonardi and R.Y. Nanetti (1993), *Making Democracy Work*, Princeton, NJ, Princeton University Press.

Rothstein, B. and E.M. Uslaner (2005), "All for all: Equality, corruption, and social trust", *World Politics, Vol.* 58, pp. 41-72.

Sangnier, M. (2010), "Does trust favor macroeconomic stability?", *PSE Working Papers,* No. 2009-40.

Sen, A. (2002), "Health: Perception versus observation", *British Medical Journal*, Vol. 324, pp. 860-861.

Senik, C. (2011), "The French unhappiness puzzle: The cultural dimension of happiness", *Paris School of Economics Working Papers*, No. 2011-34, Paris-Jourdan Sciences Économiques.

SOM Institute (2015), *The National SOM Survey* (database), *http://som.gu.se/som_institute/-surveys/national-som* (accessed 05 March 2017).

Statistics Canada (2013), *Trends in Social Capital in Canada, www.statcan.gc.ca/pub/89-652-x/89-652-x2015002-eng.htm* (accessed 10 March 2017).

Stats New Zealand (2014), *New Zealand General Social Survey*, *www.stats.govt.nz/nzgss2014* (accessed 20 April 2017).

Stolle, D. (1998), "Bowling together, bowling alone: The development of generalized trust in voluntary associations", *Political Psychology*, Vol. 19(3), pp. 497-525.

Stolle, D., S. Soroka and R. Johnston (2008), "When does diversity erode trust? Neighborhood diversity, interpersonal trust and the mediating effect of social interactions", *Political Studies*, Vol. 56, pp. 57-75.

Tabachnick, B.G. and L.S. Fidell (2001), *Using Multivariate Statistics*, Fourth edition, Needham Heights, MA: Allyn and Bacon.

Temple, J. (2000), "Growth effects of education and social capital in the OECD countries", *OECD Economic Studies*, Vol. 2001/2, OECD Publishing, Paris, *http://dx.doi.org/10.1787/eco_studies-v2001-art11-en*.

Tokuda Y. et al. (2008), "Interpersonal trust and quality-of-life: A cross-sectional study in Japan", *PLoS ONE, Vol.* 3(12).

Ubel, P.A. et al. (2005), "What is perfect health to an 85-year-old? Evidence for scale recalibration in subjective health ratings", *Medical Care*, Vol. 43(10), pp. 1054-1057.

Uslaner, E.M. (2000), "Producing and consuming trust", *Political Science Quarterly*, Vol. 115, pp. 569-590.

van de Vijver, F.J.R. and Y.H. Poortinga (1997), "Towards an integrated analysis of bias in cross-cultural assessment", *European Journal of Psychological Assessment*, Vol. 13(1), pp. 29-37.

Veenhoven, R. (2008), "The international scale interval study: Improving the comparability of responses to survey questions about happiness", *Quality of Life and the Millennium Challenge: Advances in Quality-of-Life Studies, Theory and Research, Social Indicators Research Series*, V. Moller and D. Huschka (eds.), Springer, Vol. 35, pp. 45-58.

ANNEX 1

Illustrative examples of trust measures

Interpersonal trust

Australian General Social Survey, 2014

The next few questions are about how much you trust people and institutions.

How strongly do you agree or disagree with the following statements:

That most people can be trusted?

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Strongly disagree

Canadian General Social Survey, 2013

Generally speaking, would you say that most people can be trusted or that you cannot be too careful in dealing with people?

- ☐ Most people can be trusted
- ☐ You cannot be too careful in dealing with people
- ☐ Don't know
- ☐ Refusal

On a scale of 1 to 5 (1 = cannot be trusted at all, 5 = can be trusted a lot, 88 = don't know, 99 = refused).

How much do you trust each of the following groups of people:

people in your family?	1 2 3 4 5 88 99
people in your neighbourhood?	1 2 3 4 5 88 99
people you work or go to school with?	1 2 3 4 5 88 99
strangers?	1 2 3 4 5 88 99

If you lost a wallet or purse that contained two hundred dollars, how likely is it to be returned with the money in it, if it was found:

by someone who lives close by?

by a stranger?

Read categories to respondent.

- ☐ Very likely
- ☐ Somewhat likely
- ☐ Not at all likely
- ☐ Don't know
- ☐ Refusal

European Social Survey, Wave 6

Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? *Please tell me on a score of 0 to 10 (0 = you can't be too careful, 10 = most people can be trusted, 88 = don't know).*

0 1 2 3 4 5 6 7 8 9 10 88

Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair? *(0 = most people would try to take advantage of me, 10 = most people would try to be fair, 88 = don't know).*

0 1 2 3 4 5 6 7 8 9 10 88

Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves? *(0 = people mostly look out for themselves, 10 = people mostly try to be helpful, 88 = don't know).*

0 1 2 3 4 5 6 7 8 9 10 88

European Union Statistics on Income and Living Conditions (EU-SILC), 2013

Some people say that you can trust most people. Others think you cannot be too careful in dealing with other people. Do you think most people can be trusted? *Please answer on a scale of 0 to 10, where "0" means "you cannot trust anyone" and "10" means "you can trust most people." (88 = don't know)*

0 1 2 3 4 5 6 7 8 9 10 88

European Quality of Life Survey, 2012

Generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people? *Please tell me on a scale of 1 to 10 (1 = you can't be too careful, 10 = most people can be trusted, 88=don't know, 99 = refusal).*

1 2 3 4 5 6 7 8 9 10 88 99

Falk, Becer, Dohmen, Enke, Huffman, and Sunde, 2015

How well do the following statements describe you as a person? *Please indicate your answer on a scale from 0 to 10 (0 = does not describe me at all, 10 = describes me perfectly). You can also use any numbers between 0 and 10 to indicate where you fall on the scale.*

I assume that people have only the best intentions

0 1 2 3 4 5 6 7 8 9 10

Gallup World Poll

In the city or area where you live, imagine that you lost your wallet… or something holding your identification or address… and it was found by someone else. Do you think your wallet (or your valuables) would be returned to you if it were found by a neighbour?

- ☐ Yes
- ☐ No

In the city or area where you live, imagine that you lost your wallet… or something holding your identification or address… and it was found by someone else. Do you think your wallet (or your valuables) would be returned to you if it were found by a stranger?

- ☐ Yes
- ☐ No

Mexico National Survey of Victimization and Perception of Public Security (ENVIPE), 2016

Tell me how much trust you have in… *Register the code corresponding to each choice*

1 neighbours

2 colleagues from work/business, school

3 relatives (siblings, uncles, cousins, etc.)

4 friends

- ☐ A lot…… 1
- ☐ Some…… 2
- ☐ A little…….. 3
- ☐ None…..…… 4
- ☐ Does not apply.. 5
- ☐ Don't know… 9

Naef and Schupp, 2009

In general, you can trust people.

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Somewhat disagree
- ☐ Strongly disagree

Nowadays, you can't rely on anybody.

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Somewhat disagree
- ☐ Strongly disagree

How much do you trust strangers you meet for the first time.

- ☐ No trust at all
- ☐ Little trust
- ☐ Quite a bit of trust
- ☐ A lot of trust

When dealing with strangers, it's better to be cautious before trusting them.

In general, you can trust people.

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Somewhat disagree
- ☐ Strongly disagree

New Zealand General Social Survey, 2014

And now a general question about trust *(0 = means you do not trust an institution at all, 10 = you have complete trust, 88 = don't know, 99 = refused)*.

In general how much do you trust most people in New Zealand?

0 1 2 3 4 5 6 7 8 9 10 88 99

Poland Social Cohesion Survey, 2015

Do you generally trust…?

1 people in general

2 immediate family – parents, children, spouse

3 acquaintances, friends

4 people with whom you work everyday

5 neighbours

6 strangers who you meet for the first time (*to be added in 2018)*

- ☐ I definitely trust
- ☐ I rather trust
- ☐ I rather don't trust
- ☐ I definitely don't trust
- ☐ Hard to say

UK community Life Survey, 2015-16

First, how comfortable would you be asking a neighbour to keep a set of keys to your home for emergencies, for example if you were locked out?

- ☐ Very comfortable
- ☐ Fairly comfortable
- ☐ Fairly uncomfortable
- ☐ Very uncomfortable

How comfortable would you be asking a neighbour to mind your child(ren) for half an hour?

- ☐ Very comfortable
- ☐ Fairly comfortable
- ☐ Fairly uncomfortable
- ☐ Very uncomfortable

Would you say that…

…read out…

- ☐ many of the people in your neighbourhood can be trusted
- ☐ some can be trusted
- ☐ a few can be trusted
- ☐ none of the people in your neighbourhood can be trusted
- ☐ *Spontaneous only*: Just moved here

World Values Survey, Wave 6

Generally speaking, would you say that most people can be trusted or that you need to be very careful in dealing with people? *(code one answer)*:

1. Most people can be trusted.
2. Need to be very careful.

Do you think most people would try to take advantage of you if they got a chance, or would they try to be fair?

(1 = people would try to take advantage of you, 10 = people would try to be fair) (code one number):

People would try to be fair	1	2	3	4	5	6	7	8	9	10
People would take advantage of you	1	2	3	4	5	6	7	8	9	10

I'd like to ask you how much you trust people from various groups. Could you tell me for each whether you trust people from this group completely, somewhat, not very much or not at all? *(Read out and code one answer for each): (1 = Trust completely, 2 = Trust somewhat, 3 = Do not trust very much, 4 = Do not trust at all)*.

Your family	1	2	3	4
Your neighbourhood	1	2	3	4
People you know personally	1	2	3	4
People you meet for the first time	1	2	3	4
People of another religion	1	2	3	4
People of another nationality	1	2	3	4

Institutional trust

Australian General Social Survey, 2014

The next few questions are about how much you trust people and institutions.

How strongly do you agree or disagree with the following statements:

That the healthcare system can be trusted?

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Strongly disagree

That the police can be trusted?

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Strongly disagree

That the justice system can be trusted?

- ☐ Strongly agree
- ☐ Somewhat agree
- ☐ Neither agree nor disagree
- ☐ Somewhat disagree
- ☐ Strongly disagree

Canadian General Social Survey, 2013

How much confidence do you have in the police?

Is it…?

Read categories to respondent.

Exclude security guards, fire marshals, by-law officers and all others who have no authority to make arrests.

- ☐ A great deal of confidence
- ☐ Some confidence
- ☐ Not very much confidence
- ☐ No confidence at all
- ☐ Don't know, Refuse to respond

How much confidence do you have in the Canadian Criminal courts?

Is it…?

Read categories to respondent.

- ☐ A great deal of confidence
- ☐ Some confidence
- ☐ Not very much confidence
- ☐ No confidence at all
- ☐ Don't know, Refuse to respond

If you lost a wallet or purse that contained two hundred dollars, how likely is it to be returned with the money in it, if it was found:

by a police officer?

Read categories to respondent.

- ☐ Very likely
- ☐ Somewhat likely
- ☐ Not at all likely
- ☐ Don't know
- ☐ Refusal

Eurobarometer, May 2016

I would like to ask you a question about how much trust you have in certain institutions. For each of the following institutions, please tell me if you tend to trust it or tend not to trust it.

(read out) (1 = Tend to trust, 2 = Tend not to trust, 3 = Don't know).

Justice\ the (*national*) legal system	1	2	3
The police	1	2	3
The army	1	2	3
Public administration in (*our country*)	1	2	3
Political parties	1	2	3
Regional or local public authorities	1	2	3
The (*national*) government	1	2	3
The (*national*) parliament	1	2	3
The European Union	1	2	3
The United Nations	1	2	3

European Social Survey, Wave 6

Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out *(0 = means you do not trust an institution at all, 10 = you have complete trust, 88 = don't know)*.

Firstly…read out

[*country*]'s parliament?	1	2	3	4	5	6	7	8	9	10	88
The legal system?	1	2	3	4	5	6	7	8	9	10	88
The police?	1	2	3	4	5	6	7	8	9	10	88
Politicians?	1	2	3	4	5	6	7	8	9	10	88
Political parties?	1	2	3	4	5	6	7	8	9	10	88
The European Parliament?	1	2	3	4	5	6	7	8	9	10	88
The United Nations?	1	2	3	4	5	6	7	8	9	10	88

European Union Statistics on Income and Living Conditions (EU-SILC), 2013

How much do you trust …? Please answer on a scale of 0 to 10, where "0" means "do not trust at all" and "10" means "completely trust." (88 = don't know)

The political system in [*country*]	1	2	3	4	5	6	7	8	9	10	88
The legal system in [*country*]	1	2	3	4	5	6	7	8	9	10	88
The police in [*country*]	1	2	3	4	5	6	7	8	9	10	88

European Quality of Life Survey, 2012

Please tell me how much you personally trust each of the following institutions. Please tell me on a scale of 1 to 10 *(1 = you do not trust at all, 10 = you trust completely, 11 = don't know, 12 = refusal)*.

Country parliament	1	2	3	4	5	6	7	8	9	10	88	99
The legal system	1	2	3	4	5	6	7	8	9	10	88	99

The press	1	2	3	4	5	6	7	8	9	10	88	99
The police	1	2	3	4	5	6	7	8	9	10	88	99
The government	1	2	3	4	5	6	7	8	9	10	88	99
The local (municipal) authorities	1	2	3	4	5	6	7	8	9	10	88	99

Gallup World Poll

In the city or area where you live, imagine that you lost your wallet… or something holding your identification or address… and it was found by someone else. Do you think your wallet (or your valuables) would be returned to you if it were found by the police?

- ☐ Yes
- ☐ No

Mexico National Survey of Victimization and Perception of Public Security (ENVIPE), 2016

How much trust does [authority] inspire in you?

Read the response options and register the corresponding code for each choice.

1 Traffic police of your town

2 Preventive police of your town

3 State police

4 Federal police

5 Ministerial or judicial police

6 The Public Prosecutor's Office

7 The Attorney General's Office

8 Army

9 Marine

10 Judges

- ☐ Much trust…… 1
- ☐ Some trust…… 2
- ☐ Some distrust… 3
- ☐ Much distrust… 4
- ☐ Don't know… 9

Tell me how much trust you have in prisons.

Select a single option.

- ☐ A lot…… 1
- ☐ Some…… 2
- ☐ A little…….. 3
- ☐ None…..…… 4
- ☐ Does not apply 5
- ☐ Don't know… 9

New Zealand General Social Survey, 2014

The next questions are about whether you have trust in various institutions in New Zealand.

Even if you have had very little or no contact with these institutions, please base your answer on your general impression of these institutions *(0 = not at all, 10 = completely, 88 = don't know, 99 = refused)*.

How much do you trust:

the police?	1	2	3	4	5	6	7	8	9	10	88	99
the education system?	1	2	3	4	5	6	7	8	9	10	88	99
the media?	1	2	3	4	5	6	7	8	9	10	88	99

Please remember to base your answer on your general impression of these institutions.

(0 = not at all, 10 = completely, 88 = DK, 99 = refused)

How much do you trust:

the courts?	1	2	3	4	5	6	7	8	9	10	88	99
Parliament?	1	2	3	4	5	6	7	8	9	10	88	99
the health system?	1	2	3	4	5	6	7	8	9	10	88	99

Poland Social Cohesion Survey, 2015

Do you generally trust…?

1 the army

2 Roman Catholic Church

3 the police

4 the fire brigades

5 the emergency medical services

6 the local government

7 courts

8 the national government

9 the national parliament

10 banks *(to be added in 2018)*

- ☐ I definitely trust
- ☐ I rather trust
- ☐ I rather don't trust
- ☐ I definitely don't trust
- ☐ Hard to say

World Values Survey, Wave 6

I am going to name a number of organizations. For each one, could you tell me how much confidence you have in them: is it a great deal of confidence, quite a lot of confidence, not very much confidence or none at all? *(read out and code one answer for each):*

(1 = a great deal, 2 = quite a lot, 3 = not very much, 4 = none at all).

The churches*	1	2	3	4

The armed forces	1	2	3	4
The press	1	2	3	4
Television	1	2	3	4
Labour unions	1	2	3	4
The police	1	2	3	4
The courts	1	2	3	4
The government (in your nation's capital)	1	2	3	4
Political parties	1	2	3	4
Parliament	1	2	3	4
The civil service	1	2	3	4
Universities	1	2	3	4
Major companies	1	2	3	4
Banks	1	2	3	4
Environmental organizations	1	2	3	4
Women's organisations	1	2	3	4
Charitable or humanitarian organisations	1	2	3	4
The [European Union]**	1	2	3	4
The United Nations	1	2	3	4

* [Substitute *religious organisations* in non-Christian countries; *the Church* in Catholic countries].

** [Substitute appropriate regional organisation outside Europe (e.g. in North America, NAFTA)].

ANNEX 2

Question modules

Use of the question modules

In recognition of the different users' needs and resources available to statistics producers, five question modules, rather than a single set of questions, are presented here (A to E). Each question module focuses on a distinct approach to measuring trust. Question Module A, the "core module", contains the core measures of both interpersonal and institutional trust for which international comparability is the highest priority. Module A is unique in that data producers – particularly national statistical agencies – are encouraged to implement it ***in its entirety***. When this is not possible, the single primary measure outlined in the module should be used at a minimum. Modules B to E focus on different approaches to measuring trust in greater depth. Unlike the core module, these modules are not intended to be used in their entirety or unaltered, but provide a resource for national statistical agencies that are developing their own questionnaires. Data producers can use these modules partly or combine them as needed.

A detailed discussion of the rationale behind the different questions included in each module is contained in Chapter 4. In particular, Section 4.4 includes explanations for each module, outlining the reasons for the choices made with regard to wording and scale use as well as decisions around what to include in each module. In general, all questions in this Annex 2 have been selected based on them having been used and tested in household surveys before.

Module A. Core measures

Objective

This module is intended to provide a minimal set of measures of trust that cover both interpersonal trust and institutional trust and could readily be included in household surveys. The core measures included here are the measures for which there is the strongest evidence for their validity and relevance, and for which international comparability is the most important. Data producers wishing to gather trust data are recommended in include in their survey the questions in this core module ***as they are*** shown below.

The additional modules included in this Annex (module B to D) should be seen as ***additions*** (rather than replacements) of the core module; they should be used by data producers wishing to collect more detailed trust data than those in the core module (hence some questions from the core module reappear in other modules).

Description

The module contains a single question on generalised trust (A1). This question, focusing on the aspect of trust with the widest general use and strongest validity, aims to capture the most important information on generalised trust while imposing the minimum level of respondent burden. It is envisaged that question A1 will serve as the ***primary measure*** of trust when limited resources allow for only a single measure. The complete core module also includes a second question on interpersonal trust (A2), focusing on limited trust and intended to complement A1. Finally, three questions on institutional trust are listed (A3 to A5). These questions capture two of the main underlying dimensions of institutional trust, the political system and the law and order system, as identified via factor analysis (Box 2.1). Question A5 is intended to establish whether the respondent views the non-political civil service differently to political institutions more generally. The three institutional trust questions are of a more experimental nature than the questions on interpersonal trust, but deemed of sufficient validity for data producers to include them in the core module.

Box A.2.1. **Core questions**

A1. And now a general question about trust. On a scale from zero to ten, where zero is not at all and ten is completely, in general how much do you trust most people?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

A2. On a scale from zero to ten, where zero is not at all and ten is completely, in general how much do you trust most people you know personally?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

The next questions are about whether you have trust in various institutions in [COUNTRY].

Even if you have had very little or no contact with these institutions, please base your answer on your general impression of these institutions.

Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly… **READ OUT**

A3. [Country's] Parliament?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

A4. The police?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

A5. The civil service?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

Origin

The primary question is based on the generalised trust question from the New Zealand General Social Survey 2014. This, in turn, is derived from the widely used Rosenberg question on generalised trust (Rosenberg, 1957) but focuses on *trust* rather than on *caution when dealing with people*, and is amended to use a 0-10 scale with end labels and with neutral question wording at both ends of the scale. The rationale and evidence supporting this choice of wording and response scale is outlined in Chapter 4's Section 4 as well as in Sections 3 and 4 of Chapter 3.

The additional question on limited trust is derived from the World Values Survey module on the extent of trust and has been modified to adopt consistent wording with A1.

The three institutional trust questions included here are based on wave 6 of the European Social Survey. However, only two of the seven institutions covered by the ESS module are included in the core module here (parliament and the police). These two institutions selected reflect the results of factor analysis of the underlying dimensionality of institutional trust discussed in Chapter 2 of these Guidelines (Box 2.1). The third institution (the civil service) is added from the World Values Survey. Although in the factor analysis the civil service loads, albeit with a weaker score, in the same dimension as parliament (the political system), this institution is less connected to the current government. This question will thus help to establish whether respondents view non-political and political intuitions differently.

Time

The module is expected to take about 90 seconds to complete in full. The primary question alone (A1) takes 15 to 20 seconds to complete.

Output

All five questions are collected in numerical format on a 0-10 scale. Data on interpersonal trust can be presented as the mean value of responses, excluding missing values, or as the proportion of the population reporting trust under a given threshold. On the basis of existing data, the proportion reporting a score of 0-4 might be a reasonable threshold. Standard measures of distribution should be reported, including the standard deviation of responses and the inter-quartile range of responses.

Information should be broken down by relevant population groups, ideally showing both the mean and proportion with a score of 0-4 for each population group.

None of the questions in this module should be combined into an index of "overall" trust in public institutions.

Guidelines for interviewers

The primary question (A1) is intended to be general with respect to who is meant by *most people*, and guidance on how the question should be interpreted should be avoided. If the respondent specifically asks who is meant by the term *most people*, it would be acceptable to clarify that this means *anyone in this country*.

Question A2 focuses on people that the respondents know personally. This should be taken to include not just close friends and family, but also other people with whom the respondent has an ongoing relationship. It is not intended to include people the respondent has met only once.

The three questions related to institutional trust (A3 to A5) are intended to capture the respondent's trust in the institution itself rather than the current performance of the institution, or broader issues relating to the respondent's views of the current government in power. If the respondent asks for clarification, it should be stressed that he/she is being asked to evaluate the institution (e.g. parliament) rather than the current leadership of the institution (e.g. the political party in power). Further, interviewers should not provide respondents with any kind of narrow definition of the institution asked about in the question when posing the question itself.

It is recommended that the order of the three institutions is randomised across the survey to minimise order effects.

Module B. Experimental module on evaluations

Objective

The objective of this experimental module is to assess trust levels based on an evaluation of the respondent's own feelings (trust, confidence or a related concept) at the current point in time rather than collecting information about the future or about past experiences. This module should be seen as additional to the core module. As the evidence base currently available to assess its statistical quality is weaker than in the case of the core module, the questions included below could be revised as better evidence becomes available in the future.

Description

The module has two sub-sections, the first focusing on interpersonal trust and the second on institutional trust. It is not envisaged that the module would be used in its entirety, but rather that it be used as a resource for questionnaire developers where more detail on measures of trust is required than is provided in the core module. Data producers can select and adjust questions as necessary for their needs (apart from B1 to B4, which should be used unaltered). This also explains why the response scales in this module have not been harmonised but for the most part left as they have been devised in the original source of each question. Questions B1 to B4 form a single block capturing interpersonal trust (B1 and B3) and caution when dealing with others (B2 and B4); these questions have been experimentally validated (in small samples), and are included here even if some of the issues covered overlap with those included in the core module. Questions B5 to B8 focus on collecting more detail as to who the respondent trusts, and they can be used as a full module or individually. In the case of institutional trust questions (B9 to B18), the goal is to flesh out a broader range of institutions beyond the three presented in the core module; questions on trust in the parliament, the police and the civil service (which were also included in the core module) are also included below; they can be dropped from the core module in those cases where data producers want to collect information on the broader range of institutions included in this experimental module.

Origin

Questions B1 to B4 are taken from the German Socio-Economic Panel (Naef and Schupp, 2009), and have been validated against experimental lab measures of trust. The questions relating to who the respondent trusts (B5 to B8) are derived from the World Values Survey, but the question wording and response scale have been brought into line with the 0-10 scale used in the core question module.

Box A.2.2. **Evaluative questions**

Interpersonal Trust

B1. In general, you can trust people.

Disagree strongly

Disagree somewhat

Agree somewhat

Agree strongly

B2. Nowadays, you can't rely on anybody.

Disagree strongly

Disagree somewhat

Agree somewhat

Agree strongly

B3. How much do you trust strangers you meet for the first time?

No trust at all

Little trust

Quite a bit of trust

A lot of trust

B4. When dealing with strangers, it's better to be cautious before trusting them.

Disagree strongly

Disagree somewhat

Agree somewhat

Agree strongly

Using a scale of 0 to 10 where 0 means "not at all" and 10 means "completely", how much do you trust each of the following groups of people:

B5. People in your family?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B6. People in your neighbourhood?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B7. People you work or go to school with?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B8. Strangers?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

Institutional Trust

The next questions are about whether you have trust in various institutions in [COUNTRY].

Even if you have had very little or no contact with these institutions, please base your answer on your general impression of these institutions.

Box A.2.2. **Evaluative questions** *(cont.)*

Using this card, please tell me on a score of 0-10 how much you personally trust each of the institutions I read out. 0 means you do not trust an institution at all, and 10 means you have complete trust. Firstly… **READ OUT**

B9. [Country's] Parliament?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B10. The courts?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B11. Political parties?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B12. Politicians?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B13. The police?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B14. The armed forces?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B15. The civil service?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B16. The media?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B17. The banks?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

B18. Major companies?

Not at all	Completely	DK
0 1 2 3 4 5 6 7 8 9 10		88

The institutional trust questions use the same format as questions A3 to A5, but cover a wider range of institutions. They are derived from the World Values Survey and wave 6 of the European Social Survey.

Time

Questions B1 to B4 are intended to be used together and would take approximately 60 seconds to complete. A similar amount of time is likely to be required for questions B5 to B8.

The institutional trust module is longer, and is likely to take approximately 150 seconds of survey time to complete.

Output

The four-item interpersonal trust scale comprised by questions B1 to B4 should be reported as the mean of all four questions, excluding non-responses. Distributional information can be reported based on the standard deviation of responses and inter-quartile range.

Questions B5 to B8 should be reported individually and not combined into a single scalar measure. For each measure, both the mean and the proportion with low levels of trust (0-4) should be reported, along with the standard distributional measures.

Results from the institutional trust questions should be reported individually as the mean response excluding non-responses and the proportion of the population reporting low trust (0-4). These questions should not be combined into an index of "overall" trust in institutions.

Guidelines for interviewers

These questions go into more detail with respect to trust than the questions in the core module, but are otherwise very similar.

In general, respondents should be allowed to interpret questions B1 to B4 for themselves. The question content overlaps significantly between the four questions in order to obtain a more reliable overall measure of trust. If respondents comment on this, it should be stressed that different people will interpret each question slightly differently and that the repetition across four questions is to enable a more consistent overall picture to be drawn.

Questions B5 to B18 all require a more precise response from respondents in that each question focuses on trust in a specific group or institution. The groups mentioned by each question are intended to be mutually exclusive. If respondents ask, for example, *strangers* can be taken to mean people who are not family, neighbours or work colleagues. In the case of questions B9 to B18, respondents may be concerned that they have had no interaction with a particular institution. In this case, it should be emphasised that the question is focused on ***trust*** in the institution, rather than ***satisfaction*** with the services that these institutions provide. Further, interviewers should not provide respondents with any kind of narrow definition of the institution asked about in the question when posing the question itself.

It is recommended that the order of institutions is randomised across the survey to minimise order effects.

Module C. Experimental module on expectations

Objective

This module contains questions that ask about the respondent's expectations as to what would or will happen in a given situation. These questions focus directly on the conduct expected of a third party, for example neighbours, police officers or strangers. As such, it provides measurement on the trustworthiness of a given institution, community or society. As the evidence base currently available to assess its statistical quality is weaker than in the case of the core module, the questions included below could be revised as better evidence becomes available in the future.

Description

The module has two sub-sections, the first focusing on interpersonal trust and the second on institutional trust. It is not envisaged that the module would be used in its entirety, but rather that it would be used as a resource for questionnaire developers where more detail on measures of trust is required than is provided in the core module. Data producers can select and adjust questions as necessary for their needs. In the case of institutional trust, the questions have been designed under the five dimensions that have been identified by the OECD Trust Framework as drivers of trust in public institutions.

Box A.2.3. **Expectations questions**

Interpersonal Trust

In the city or area where you live, imagine that you lost your wallet or something holding your identification or address and it was found by someone else.

C1. Do you think your wallet (or your valuables) would be returned to you if it were found by a neighbour?

1. Yes
2. No

C2. Do you think your wallet (or your valuables) would be returned to you if it were found by a stranger?

1. Yes
2. No

Institutional Trust

The following questions are about your expectations of behaviour from public institutions. In each question, you will be asked whether you think a particular example of behaviour is something that would be expected not to occur at all, or to always occur.

Please respond on a scale from 0 to 10 where 0 means very unlikely and 10 means very likely.

C3. If you were to complain about bad quality of a public service, how likely is that the problem would be easily resolved?

Very unlikely	Very likely	DK
0 1 2 3 4 5 6 7 8 9 10		88

C4. If a natural disaster occurs, do you think that the provision by government of adequate food, shelter and clothing will be timely and efficient?

Very unlikely	Very likely	DK
0 1 2 3 4 5 6 7 8 9 10		88

C5. If a decision affecting your community were to be taken by the local or regional government, how likely is it that you and others in the community would have an opportunity to voice your concerns?

Very unlikely	Very likely	DK
0 1 2 3 4 5 6 7 8 9 10		88

C6. If an individual belongs to a minority group (e.g. sexual, racial/ethnic and/or based on national origin), how likely is it that the individual will be treated the same as other citizens by a government agency?

Very unlikely	Very likely	DK
0 1 2 3 4 5 6 7 8 9 10		88

Box A.2.3. **Expectations questions** *(cont.)*

C7. If taxes were to be increased, do you think that the financial burden would be shared fairly across social groups?

Very unlikely	Very likely	DK
0 1 2 3 4 5 6 7 8 9 10		88

The response scale for the following questions has changed. Please respond on a scale from 0 to 10, but 0 now means most likely to refuse and 10 most likely to accept.

C8. If a private citizen offers a government employee an improper payment in order to speed up administrative procedures, do you think that the government employee would accept the bribe?

Most likely to refuse	Most likely to accept	DK
0 1 2 3 4 5 6 7 8 9 10		88

C9. If a large business offered a well-paid job to a high-level politician in exchange for political favours during their time in office, do you think that the politician would accept this proposal?

Most likely to refuse	Most likely to accept	DK
0 1 2 3 4 5 6 7 8 9 10		88

C10. If a member of [country's parliament] were offered a bribe to influence the awarding of a government contract, do you think that the member of parliament would accept the bribe?

Most likely to refuse	Most likely to accept	DK
0 1 2 3 4 5 6 7 8 9 10		88

Origin

The questions on interpersonal trust included here are based on the Gallup World Poll with some minor wording changes based on the conclusions of Chapter 2. Questions C3 to C10 were developed by the OECD to measure the main drivers of trust in the OECD Trust Framework (OECD, 2017). These questions were also included in the pilot countries of the *Trustlab* project.

Time

Questions C1 and C2 can be expected to be completed in approximately 30 seconds. The OECD question module on institutional trust is longer, and might be expected to take approximately 3 minutes of survey time to complete.

Output

Questions C1 and C2 should be presented as the proportion of the population answering *yes* to each of the questions. Because the questions are intended to capture the level of trust that the respondent has in different types of people, the questions should not normally be combined into a single index.

Questions C3 to C10 are intended to be used as a complete module and capture the five main drivers of institutional trust from the OECD Trust Framework (responsiveness, openness, reliability, fairness and integrity) under the umbrella concepts of a government's competence and values; question C3 relates to responsiveness; C4 to reliability; C5 to openness; C6 and 7 to fairness; and C8 to C10 to integrity. The responses to the questions can be reported as the mean value excluding non-responses, and in the cases where more than one question per driver is

included as index calculated as the mean response for the questions relating to the relevant dimension of trust, omitting non-responses. For example, the index value for integrity could be computed as the mean value of responses to questions C8, C9 and C10.

In addition to the indices on the drivers of trust, it is also possible to group the questions into the two broad dimensions of competence (C3 to C4) and values (C5 to C10). These indices combine the drivers into the two dimensions that are thought to influence how people form views of the trustworthiness of various institutions. Both indices can be calculated as the mean of the relevant sub-indexes/questions, excluding non-responses.

Guidelines for interviewers

Expectations questions such as those included in this module require the respondent to provide their views on the likely outcome of a hypothetical situation. One possible concern from respondents may be that they have no experience of the kind of situation discussed. This is particularly the case for the institutional trust questions, where a respondent might be concerned that they have never had contact with the institution in question. In all cases, it is important to emphasise that the information that this module is designed to collect is not an estimate of how likely the situations in question actually are, but rather of what the respondent believes. While it is very likely that respondent beliefs will differ significantly from the actual likelihood of many situations occurring, it is the respondent's beliefs that will drive their attitudes and behaviours. For this reason, respondents should be encouraged to respond with what they think will happen, even if they have no actual experience relevant to the situation in question.

Module D. Experimental module on experiences

Objective

This module collects information based on the respondent's past experiences and behaviour rather than their values, perceptions or evaluations. In consequence, respondents are confronted with a series of questions related to situations that they may have experienced and that are typically associated with a trusting behaviour. Questions of this type have been used to validate other measures of trust and are useful to confirm the validity of other sorts of trust measure. It is not envisaged that the module would be used in its entirety, but rather that it would be used as a resource for questionnaire developers where more detail on measures of trust is required than is provided in the core module. Data producers can select and adjust questions as necessary for their needs. This also explains why the response scales in this module have not been harmonised but for the most part left as they have been devised in the original source of each question. As the evidence base currently available to assess its statistical quality is weaker than in the case of the core module, the questions included below could be revised as better evidence becomes available in the future.

Description

The module has two sub-sections, the first focusing on interpersonal trust and the second on institutional trust. The module on interpersonal trust focuses on the respondent's past behaviour and identifies the degree to which the respondent has recently behaved in a way that is consistent with trusting behaviour. In the case of institutional trust, the questions relate to behaviours that reflect confidence in that specific institution. There are three questions relating to interpersonal trust and two relating to institutional trust.

Box A.2.4. **Experiential questions**

Interpersonal Trust

D1. How often do you lend personal possessions to your friends?

1. Never
2. Infrequently
3. Sometimes
4. Often
5. Very often

D2. How often do you lend money to your friends?

1. Never
2. Infrequently
3. Sometimes
4. Often
5. Very often

D3. How often do you leave your door unlooked?

1. Never
2. Infrequently
3. Sometimes
4. Often
5. Very often

Institutional Trust

Have you done any of the following in the past month?

How about:

D4. Voiced your opinion to a public official?

Yes

No

D5. Signed a petition?

Yes

No

Origin

The questions on interpersonal trust are based on a set of questions used by Naef and Schupp (2009) to validate a survey module on trust in the German Socio-Economic Panel. The two questions on institutional trust are taken from the Gallup World Poll.

Time

This module can be expected to take approximately 60 to 80 seconds to complete.

Output

The questions on interpersonal trust are collected as a five-point labelled Likert scale. For output purposes, they can be reported as the proportion of respondents in each output

category, or alternatively some of the output categories can be grouped together as a single category. In addition, it is useful to report the mean numeric value for each question, excluding non-responses. Because all three questions relate to the same underlying concept, it is also possible to construct an index of trusting behaviour calculated as the mean value of questions D1 to D3, excluding non-responses. Standard measures of distribution used should be the standard deviation of responses and the inter-quartile range of responses. The mean value of responses, and the standard error of this estimate, could be used to describe differences in responses to the various questions among sub-groups of the population.

The questions relating to institutional trust should be reported as the percentage of respondents replying *yes* to the question.

Module E. Experiments

Objective

One recent alternative to collecting information on trust through surveys involves experimental techniques to measure trusting and trustworthy behaviour by respondents in controlled conditions. Pioneered by academic psychologists and experimental economists, games such as the Trust Game and the Dictator Game collect information on the respondent's actual behaviour in circumstances where genuine rewards – typically, but not always, relatively small amounts of money – are at stake.

Description

The experimental module described here includes two games, the protocols for which are detailed in Box A.2.5. The first of these is the Dictator Game, which provides an experimental measure of altruism. The Dictator Game does not itself provide information on trust, but information on altruism is important in order to control for the impact of altruism on the first mover's choice in the Trust Game. The second game included here is the Trust Game, which provides an experimental measure of trust and a measure of trustworthiness. Both games are adapted for use in internationally comparative studies (using online formats), such as the OECD's *Trustlab* project.

The Trust and Dictator Games can be designed in different variations, depending on the mode of implementation and the exact research purpose. In in-person games, such as in Berg et al. (1995), respondents interact with other participants live, meaning a responder learns the actions of the first mover and then chooses a response. This process is also known as the direct-response method. For practical reasons, this is difficult when implementing an online module such as *Trustlab*, because participants would need to log on to the platform at exactly the same time. To circumvent this, the module below relies on the so-called strategy method (Selten, 1967), in which a responder makes conditional decisions for each possible information set (the strategy method thus has the added advantage of collecting data for each information set, including nodes that are only reached occasionally in the actual course of play). The strategy method also facilitates matching participants with players who have previously logged on to the platform – players hence find out about their payoff based on *ex post* matching at the end of the study. A review by Brandts and Charness (2011) established that most studies found no difference in experimental results between the direct-response and strategy methods.

Traditionally, participants are randomised into one role (either Participant A or Participant B) in each game – as described in the protocols in Box A.2.5. In *Trustlab*,

participants in the Trust Game are asked to play, in turn, as both Participant A and Participant B. This way, a measure of both trust and trustworthiness is obtained for each respondent. The choice to randomise roles or have participants play both consecutively involves a trade-off between saving data collection time (the average time per participant increases when asked to play both roles) and doubling the number of observations. The impact of playing both roles as opposed to being randomised into only one role is documented by Burks et al. (2003), who found that playing both roles somewhat reduces trust and trustworthiness, but only when the participant is informed of this set-up in advance.

Participants are typically rewarded with real monetary incentives in order for the Dictator Game and Trust Game to raise the stakes and mimic real-world behaviour. In the context of OECD countries, starting endowments equivalent to USD 10 (PPP) are conventional and commonly considered a sufficient incentive to incite real behaviour (Johnson and Mislin, 2011). Increasing endowments has been found to influence Participant Behaviour, with higher stakes associated with less trust and less altruistic behaviour (Engel, 2011; Johnson and Mislin, 2011). If multiple experimental games are played, rewards can be randomised between games in order to economise the study design (Davis and Holt, 1993; Johnson and Mislin, 2011). Further, if multiple games are played, it is recommended to randomise their order.

The module below elicits responses focused on interpersonal interactions within a country. These protocols can be modified to focus on interpersonal trust within or between specific groups. By changing the description of the participant's characteristics, these games can be used to measure trust across age, gender, income or ethnic group, among others (Lei and Vesely, 2010; Stanley et al., 2012).

Box A.2.5. **Experimental measures of trust**

General Preface

At the beginning of each task, you will be grouped with other study participants, who are respondents from [country name] like you.

In each task, the other participants you interact with will be different: you will never interact with the same person more than once. Note that the other participants may not be online at this exact moment – we will record your and their answers and will match you at the end of the study.

Your earnings in each task will depend on your and the other participant's decisions.

At the end of the study, one of the tasks you have completed will be randomly selected. The amount of money you will receive will correspond to your earnings in this selected task.

Dictator Game Protocol

This task involves two participants: Participant A and Participant B.

These are the rules:

At the beginning of the task, Participant A receives [equivalent of USD 10].

Participant B does not receive any endowment: he or she has zero money.

Participant A must now decide whether he or she wants to transfer any of his or her [equivalent of USD 10] to Participant B.

Participant B cannot transfer any amount back to Participant A.

If selected as Participant A: You are selected to be in the role of Participant A. You have [equivalent of USD 10] in your possession.

Box A.2.5. **Experimental measures of trust** *(cont.)*

How much (if anything at all) do you want to transfer to Participant B?

If selected as Participant B: You are selected to be in the role of Participant B. This is a passive role.

Trust Game Protocol

This task involves two participants: Participant A and Participant B.

These are the rules:

At the beginning of the task, both participants receive [equivalent of USD 10].

Participant A has the option to transfer none, part or all of his or her [equivalent of USD 10] to Participant B.

Whatever amount Participant A transfers to Participant B is **multiplied by 3**.

Participant B, after receiving the transfer of Participant A, has to decide how much (if any) money he or she wants to send back to Participant A.

If selected as Participant A: You are selected to be in the role of Participant A. You have [equivalent of USD 10] in your possession.

How much (if anything at all) do you want to send to Participant B?

If selected as Participant B: You are selected to be in the role of Participant B. You have [equivalent of USD 10] in your possession. Remember that Participant A also starts with an endowment of [equivalent of USD 10].

If Participant A sends you 0 (your total endowment is [equivalent of USD 10]), how much (if anything at all) will you send back to Participant A?

[…]

If Participant A sends you [equivalent of USD 10] (your total endowment is [equivalent of USD 40]), how much (if anything at all) will you send back to Participant A?

Origin

The Trust game is derived from Berg, Dickhaut and McCabe (1995). Since the introduction of this game (commonly referred to as the BDM game, after its three authors), it has been repeated extensively. A thorough review of 162 studies involving the Trust Game can be found in Johnson and Mislin (2011). The original version of the Dictator Game was first proposed by Kahneman et al. (1986), and similar to the Trust Game, has been repeated extensively across different contexts (see Engel, 2011, for a meta-analysis).

Time

The Trust Game and the Dictator Game each take about 5 minutes to complete.

Output

The Dictator Game provides a measure of *altruism* by recording the amount that a participant is willing to transfer to an anonymous other participant, without any form of reciprocation. Since a participant has the choice to send integers between 0 and 10 (and the total starting endowment is the equivalent of USD 10), the scale can be easily compared with survey measures of altruism, and transfers as a proportion of the initial endowment can be intuitively derived. In the Trust Game, two outputs are collected: *trust and trustworthiness*.

Trust is equal to the proportion of his or her initial endowment that Participant A is willing to send to Participant B: it is a manifestation of the belief that Participant B will engage in co-operative behaviour and reward Participant A for his or her willingness to do so. Trustworthiness is equal to the amount that Participant B sends back to Participant A, conditional on the amount received. It is hence a manifestation of the degree to which Participant B is willing to reciprocate Participant A's co-operative behaviour (Camerer, 2003).

Guidelines for interviewers

There are several important procedural factors that need to be explained clearly to respondents. First, it is important to state clearly *whom* participants are interacting with in both the Dictator and Trust Games. In a nationally representative study aimed at measuring trust on a national level, the participant should be matched with a random other respondent from the same country. This should be clarified before explaining the game. When multiple games are played, it is best practice to remind participants that the other participants are also from the same country, and that the participants are not the same as in the previous game, emphasising the non-repeated nature of the games. It is also important that participants are aware that they are interacting in the game in an anonymous manner. This prevents fear of retribution or strategic behaviour (Johnson and Mislin, 2011). In online data collections, player anonymity is implicit due to the mode of participation: participants are geographically separated, and their identity is not revealed.

The second procedural factor that needs to be highlighted is *what* the participant is playing for. The incentives and endowments need to be clearly defined and easily accessible for the respondent in order to facilitate an informed decision. There are two approaches to representing the value of stakes in the Trust and Dictator Games. In the original Berg et al. experiment (1995), physical dollar bills were used in interactions, a strategy that has been modified to simplify implementation in online games by using references to currency values, which has been the practice in many subsequent studies (Johnson and Mislin, 2011). A different approach is to create an imaginary currency (tokens) and use a consistent exchange rate (e.g. 1 token = USD 0.20 in Kim et al., 2016). This approach ensures that the incentive is of exactly the same size in different countries, but it increases cognitive burden, because the token still needs to be converted into real money with varying exchange rates. In *Trustlab*, real national currency values are used in order to remind respondents that games involve real stakes.

Finally, it is critical that participants have a full understanding of the games before making their final decisions. The protocol included here sets out the rules of the game, but it is conventional that an explanation of the rules is complemented with additional guidance, for example, in the form of typical interactions (Berg et al., 1995), a visual simulation or a practice round.

In general, instructions or materials that actively mislead participants by stating or strongly implying something that is not true (e.g. telling participants that they are playing games versus another subject when they are actually playing a confederate of the experimenter, or a computerised robot, paying participants based on something other than the announced rules, or resolving random outcomes in a manner inconsistent with announced rules) count as deception and should be avoided at all costs.

References

Berg, J., J. Dickhaut and K. McCabe (1995), "Trust, reciprocity, and social history", *Games and Economic Behavior*, Vol. 10, pp. 122-142.

Brandts, J. and G. Charness (2011), "The strategy versus the direct-response method: A first survey of experimental comparisons", *Experimental Economics*, Vol. 14, No. 3, pp. 375-398.

Burks, S., J. Carpenter and E. Verhoogen (2003), "Playing both roles in the trust game", *Journal of Economic Behavior & Organization*, Vol. 51, pp. 195-216.

Camerer, C. (2003), *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton University Press, Princeton, NJ.

Davis, D. and C. Holt (1993), *Experimental Economics*, Princeton University Press, Princeton, NJ.

Engel, C. (2011), "Dictator games: A meta study", *Experimental Economics*, Vol. 14, No. 4, pp. 583-610.

Johnson, N.D. and A.D. Mislin (2011), "Trust games: A meta-analysis", *Journal of Economic Psychology*, Vol. 32, pp. 865-889.

Kahneman, D., J.L.Knetsch, and R.H. Thaler (1986-), "Fairness and the Assumptions of Economics", *The Journal of Business*, 59 (4): pp. 285-300.

Kim, J., L. Putterman and X. Zhang (2016), "Trust and Cooperation: An experimental study", *Brown University Working Paper*, *www.brown.edu/academics/economics/candidates/sites/brown.edu.academics.economics.candidates/files/Trust%20and%20Cooperation_Kim%20et%20al..pdf*.

Lei, V. and F. Vesely (2010), "In-group versus out-group trust: The impact of income inequality", *Southern Economic Journal*, Vol. 76, No. 4, pp. 1049-1063.

Naef, M. and J. Schupp (2009), "Measuring trust: Experiments and surveys in contrast and combination", *IZA Discussion Paper,* No. 4087.

OECD (2017), *Trust and Public Policy: How Better Governance Can Help Rebuild Public Trust*, OECD Publishing, Paris, *http://dx.doi.org/10.1787/9789264268920-en*.

Rosenberg, M. (1957), "Misanthropy and attitudes toward international affairs", *Journal of Conflict Resolution*, pp. 340-345.

Selten, R. (1967), "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments", in *Beiträge zur experimentellen Wirtschaftsforschung*, Heinz Sauermann (ed.), Vol. I, Tübingen: J.C.B. Mohr (Siebeck), pp. 136-168.

Stanley, D.A., P. Sokol-Hessner, D.S. Fareri, M.T. Perino, M.R. Delgado, M.R. Banaji and E.A. Phelps (2012), "Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions", *Philosophical Transactions of the Royal Society B: Biological Sciences*, Vol. 367, No. 1589, pp. 744-753.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

The OECD is a unique forum where governments work together to address the economic, social and environmental challenges of globalisation. The OECD is also at the forefront of efforts to understand and to help governments respond to new developments and concerns, such as corporate governance, the information economy and the challenges of an ageing population. The Organisation provides a setting where governments can compare policy experiences, seek answers to common problems, identify good practice and work to co-ordinate domestic and international policies.

The OECD member countries are: Australia, Austria, Belgium, Canada, Chile, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Japan, Korea, Latvia, Luxembourg, Mexico, the Netherlands, New Zealand, Norway, Poland, Portugal, the Slovak Republic, Slovenia, Spain, Sweden, Switzerland, Turkey, the United Kingdom and the United States. The European Union takes part in the work of the OECD.

OECD Publishing disseminates widely the results of the Organisation's statistics gathering and research on economic, social and environmental issues, as well as the conventions, guidelines and standards agreed by its members.

OECD PUBLISHING, 2, rue André-Pascal, 75775 PARIS CEDEX 16
(30 2017 03 1 P) ISBN 978-92-64-27820-2 – 2017

OECD Guidelines on Measuring Trust

Trust, both interpersonal trust, and trust in institutions, is a key ingredient of growth, societal well-being and governance. As a first step to improving existing measures of trust, the *OECD Guidelines on Measuring Trust* provide international recommendations on collecting, publishing, and analysing trust data to encourage their use by National Statistical Offices (NSOs). The Guidelines also outline why measures of trust are relevant for monitoring and policy making, and why NSOs have a critical role in enhancing the usefulness of existing trust measures. Besides looking at the statistical quality of trust measures, best approaches for measuring trust in a reliable and consistent way and guidance for reporting, interpretation and analysis are provided. A number of prototype survey modules that national and international agencies can use in their household surveys are included.

These Guidelines have been produced as part of the OECD Better Life Initiative, a pioneering project launched in 2011, with the objective to measure society's progress across eleven domains of well-being. They complement a series of similar measurement guidelines on subjective well-being, micro statistics on household wealth, integrated analysis of the distribution on household income, consumption and wealth, as well as the quality of the working environment.

Consult this publication on line at ***http://dx.doi.org/9789264278219-en***.

This work is published on the OECD iLibrary, which gathers all OECD books, periodicals and statistical databases. Visit ***www.oecd-ilibrary.org*** for more information.

OECD*publishing*
www.oecd.org/publishing

THE LONDON BOOK FAIR WINNER
INTERNATIONAL EXCELLENCE Awards 2017
IN PARTNERSHIP WITH THE PUBLISHERS ASSOCIATION

ISBN 978-92-64-27820-2
30 2017 03 1 P

9 789264 278202