Organisation for Economic Co-operation and Development

STD/DOC(2017)10

Unclassified **English - Or. English**

17 October 2017

STATISTICS DIRECTORATE

Measures of interpersonal trust: Evidence on their cross-national validity and reliability based on surveys and experimental data

WORKING PAPER No.86

Lara Fleischer, Statistics Directorate, +33 1 45 24 17 65; lara.fleischer@oecd.org

JT03420907

Measures of Interpersonal Trust: Evidence on their Cross-National Validity and Reliability based on Surveys and Experimental Data

Ryan E. Carlin, Georgia State University, rcarlin@gsu.edu

Gregory J. Love, University of Mississippi, gjlove@olemiss.edu

Conal Smith,[1] conalinparis@gmail.com

[1] Conal Smith was a member of the OECD staff in the Statistics Directorate during most of the time when undertaking the work featuring in this paper.

OECD STATISTICS WORKING PAPER SERIES

The OECD Statistics Working Paper Series – managed by the OECD Statistics Directorate – is designed to make available in a timely fashion and to a wider readership selected studies prepared by OECD staff or by outside consultants working on OECD projects. The papers included are of a technical, methodological or statistical policy nature and relate to statistical work relevant to the Organisation. The Working Papers are generally available only in their original language – English or French – with a summary in the other.

OECD Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed are those of the authors.

Working Papers describe preliminary results or research in progress by the authors and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcomed, and may be sent to the Statistics Directorate, OECD, 2 rue André Pascal, 75775 Paris Cedex 16, France.

This document, as well as any statistical 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.

The release of this working paper has been authorised by Martine Durand, OECD Chief Statistician and Director of the OECD Statistics Directorate.

www.oecd.org/std/publicationsdocuments/workingpapers/

Abstract / Résumé

Interpersonal trust (i.e. trust in other people) is an issue of high interest to both policy-makers and researchers seeking to understand what drives social and economic outcomes. However, for trust to usefully inform policy and analysis it is necessary to have valid and reliable measures of it. Despite a large body of evidence on the relationship between trust and other social and economic outcomes, evidence on the validity of trust from experimental data is conflicting. In particular, while many studies find no correlation between survey measures of trust and experimental measures at an individual level, other studies suggest a significant, if modest, correlation at the country level. This article examines the relationship between survey and experimental measures of trust in others using a large dataset containing aggregate experimental and survey measures of trust from 167 studies conducted in 36 countries. Importantly, the dataset also includes individual measures of both survey and behavioural trust in seven countries, and data from two panel studies with repeated survey measures of trust. Using these multiple data sources, the paper investigates the degree to which survey measures of interpersonal trust are valid at both an individual and cross-country level. The paper shows the existence of a significant correlation between survey and experimental measures of interpersonal trust at the country-level. Evidence on measurement errors in existing small-scale studies underscores the importance of developing better quality data from both surveys and experiments.

Keywords: Interpersonal trust, trust game, measurement

JEL Classification: C83, C91, Z10

****

La question de la confiance interpersonnelle (c'est-à-dire la confiance envers les autres) revêt un intérêt particulier à la fois pour les responsables de l'action publique et pour les chercheurs qui étudient les déterminants des résultats sociaux et économiques. Néanmoins, les indicateurs de la confiance doivent être robustes et fiables pour pouvoir étayer efficacement les politiques publiques et l'analyse. S'il existe de nombreux éléments attestant de la corrélation entre la confiance et d'autres résultats sociaux et économiques, les informations concernant la validité des indicateurs de confiance issus de données expérimentales sont contradictoires. Ainsi, de nombreuses études concluent à une absence de corrélation entre les indicateurs de la confiance issus d'enquêtes et les indicateurs individuels fondés sur des données expérimentales. D'autres en revanche mettent en évidence une corrélation significative, même si elle est modeste, à l'échelon national. Le document examine la relation entre les indicateurs de la confiance intepersonnelle issus d'enquêtes et les indicateurs expérimentaux, en se fondant sur une vaste base de données contenant des indicateurs expérimentaux et issus d'enquêtes provenant de 167 études réalisées dans 36 pays. Surtout, cette base de données comprend également des indicateurs individuels de la confiance issus d'enquêtes et d'expérimentations couvrant sept pays, ainsi que des données provenant de deux études par panel fondées sur des indicateurs de la confiance issus d'enquêtes régulières. À partir de ces différentes sources de données, le document examine la validité des indicateurs de la confiance interpersonnelle issus d'enquêtes tant au niveau individuel qu'international.

Le document montre l'existence d'une corrélation significative entre les indicateurs de la confiance issus d'enquêtes et de ceux fondés sur des données expérimentales à l'échelon national. Des erreurs de mesure avérées au niveau des études utilisant de petits échantillons soulignent l'importance de développer des données de meilleure qualité provenant aussi bien des enquêtes que des données expérimentales.

Mots clés : confiance interpersonnelle, jeux portant sur la confiance, mesure

Classification JEL : C83, C91, Z10

Table of contents

Tables

Figures

Introduction

1. Trust in others matters – evidence from a range of sources demonstrates reliable associations between survey measures of interpersonal trust (i.e. trust in other people) and other social, economic, and political outcomes. For example, interpersonal trust is a prime candidate for the social capital element of unexplained total factor productivity (World Bank, 2006; Algan and Cahuc, 2013), and is correlated with affiliated outcomes such as health (Kawachi and Berkman, 2000; Boreham et al., 2002; Ginn and Arber, 2004; Stafford et al., 2004), crime (Sampson, 2012) and life satisfaction (Algan and Cahuc, 2013; Boarini et al., 2012; Helliwell and Wang, 2010). Importantly, country-average trust correlates with the individual-level outcomes that theory would predict, such as life satisfaction (Boarini et al., 2012). This broad cross-national evidence can, in part, be taken as evidence of construct validity, i.e. survey measures of interpersonal trust consistently correlate with measures tapping into the attitudes and behaviours that should, theoretically, be influenced by it.

2. In contrast to the large body of evidence for construct validity, evidence of convergent validity between survey-based measures of interpersonal trust and behavioural measures of it is quite small. This is made all the more conspicuous by the recent proliferation of studies measuring trust with experimental behavioural techniques. Pioneered by experimental and behavioural economists and more recently by psychologists and neuroscientists interested in the cognitive processes underlying decision-making, exercises such as the trust game first introduced by Berg et al. in 1995[2] reveal respondents' actual behaviour in circumstances where genuine rewards – typically, but not always, relatively small (yet large enough to form genuine incentives) – are at stake. 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. This makes the question of whether experimental measures of trust converge with survey measures of high interest.

3. Without a more definitive answer to this question, it is impossible to establish whether survey-based measures trust correlate with other important outcomes because they actually capture trust or whether they tap an unobserved variable correlated to both the survey measure of trust and the outcomes of interest. If such measures are to inform policy this is clearly problematic. In particular, better evidence on the convergent validity of survey measures of trust is necessary both to provide assurance that policy conclusions drawn from the analysis of trust measures are valid, but also to encourage national statistical offices to begin or continue collecting trust data.

4. The warrant for exploring convergent validity is made plain in the comparable case of subjective well-being measures. Determining convergent validity was central to establishing their credibility (OECD, 2013). Specifically, the robust relationship between

[2] 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; cooperation 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.

survey measures of subjective well-being and bio-physical measures was instrumental in addressing many of the concerns raised by sceptics. Unfortunately there is less bio-physical evidence available to establish the construct validity of survey measures of interpersonal trust. Fehr (2009) 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 the experimental design explicitly rejects the hypothesis that oxytocin affected the amount players were prepared to trust others by affecting their general altruism or their risk preferences, suggesting instead that the measures of trust produced by the trust game capture genuine trusting behaviour and are strongly grounded in a neuro-physical mechanism. Scans from fMRI taken during the trust game link Player 1's actions to activity in the aMPFC (Krueger et al., 2007), the area of the brain thought to be responsible for processing strategic interactions and integrating perspectives of self and other (Harris et al., 2007). In addition, Haas et al. (2015) provide evidence from fMRI scans of a sample of 82 individuals that the tendency to trust is associated with differences in the physical structure of the brain. Increased grey matter volume within area of the pre-frontal cortex and bilateral anterior insula is linked to both responses to survey questions on trust and with behavioural measures from laboratory tasks.

5. Our limited knowledge of the biophysical determinants of trust and scant work on the realm of convergent validity compounds the problem of nuanced findings in the few existing works that do examine the association between survey- and behavioural-based trusts. One consistent finding in this literature is the low correlation between behavioural measures of trust and survey measures of trust at the individual level (e.g. Glaeser et al., 2000; Lazzarini et al., 2004). A low correlation at the individual level is not particularly problematic with respect to the validity of interpersonal trust measures, which primarily relate to an individual's judgment about society as a whole. Although the analysis of survey measures shows a correlation between surveyed trust and other outcomes such as life satisfaction at the individual-level (Helliwell and Wang, 2010), this is of limited policy interest both due to endogeneity issues (i.e. direction of causality is unclear) and the potential for inflated estimates of the correlation due to shared method variance. However, the literature also shows that average trust at the *country* level predicts *individual* outcomes, which can neither be due to shared method variance nor endogenous to the respondent (Boarini et al., 2012). It is also at the aggregate level (neighbourhood, region, country, etc.) that policy interventions seeking to affect trust are most relevant. It would, therefore, be problematic if there was no cross-country correlation between behavioural measures of trust and survey measures.

6. Currently very few works look at the relationship between behavioural data on trust and survey data in a cross-country context. Johnson and Mislin (2011, 2012), conduct a meta-analysis of experimental studies involving the trust game, covering 151 replications of the Berg and McCabe trust experiment across 35 countries and over 23 000 respondents. Although most of these studies are individually small (the average sample size is 148), they cover a wide range of both developing countries (e.g. Cameroon and Uganda) and developed countries (e.g. the United States and Sweden). Contrary to earlier experimental studies focused at the individual level, Johnson and Mislin find a significant positive correlation between the World Values Survey measure of trust and trusting behaviour in experimental games.

7. This article extends Johnson and Mislin's contribution in two ways. First, it updates and expands on their already large dataset to contain aggregate behavioural and survey measures of trust from 167 studies conducted in 36 countries. This was possible by incorporating new trust game results and by the use of European Social Survey data when World Values Survey data was unavailable (both include the same trust instrument). Second, whereas Johnson and Mislin's analysis and dataset only include aggregate (between-country) results, we employ two additional datasets. Parts of the analyses presented below are based on data that include individual measures of both survey and behavioural trust in seven countries. Additionally we use data from two panel studies with repeated, within-subject survey measures of trust (in Spain and Colombia) to examine measurement reliability. Using these multiple data sources, the paper investigates the degree to which survey trust measures are valid at both an individual and cross-country level. As such, this study provides a broad assessment of the convergent validity of survey and behavioural measures of trust.

8. The paper is in four parts. Part 1 describes the datasets. Part 2 examines intra- and inter-country correlations between experimental and survey measures of interpersonal trust as well as cross-country correlations between both types of trust measure and GDP per capita. Part 3 uses panel data to estimate the measurement reliability of trust questions and compares the results to cross-national comparisons; it also looks at the relationship between non-anonymous survey questions and game play. Finally, Part 4 discusses the implications for measurement and for the policy use of trust measures.

1. Data

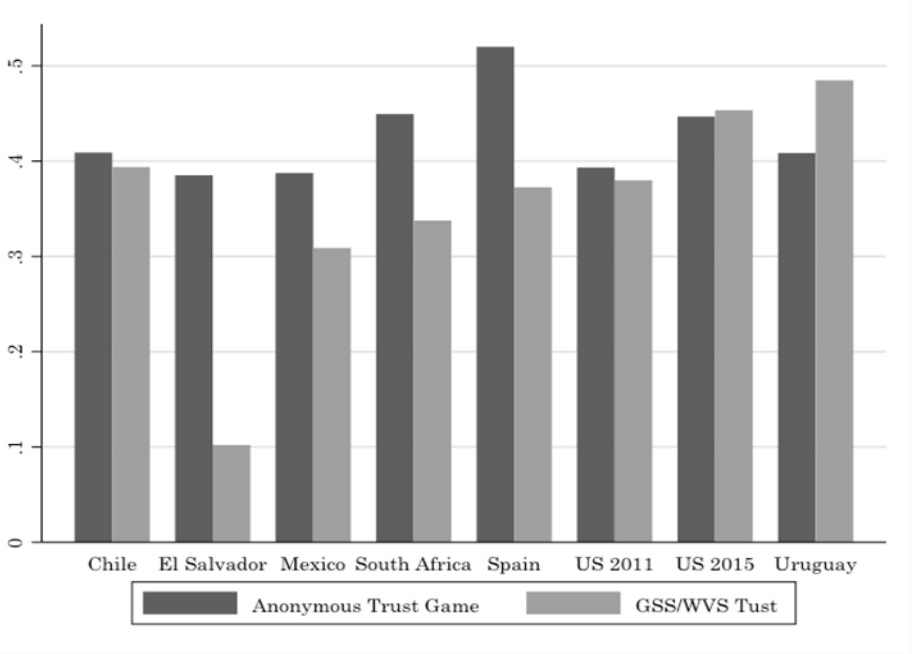
9. To assess the relationship between survey and behavioural measures of trust we employ one of the largest cross-national collections of individual-level data on survey and game-based trust. The dataset includes 2 375 subjects who both answered a survey question about interpersonal trust and participated as trustors (i.e. player 1) in trust game studies in eight studies conducted in seven countries (United States, Mexico, Chile, Uruguay, El Salvador, Spain, South Africa; see Carlin and Love (forthcoming); Martini and Torcal (forthcoming). Our survey measure is the classic interpersonal trust question originally developed for the General Social Survey (GSS) and implemented cross-nationally in the World Values Survey (WVS) often referred to as the Rosenberg question (Rosenberg, 1957). It asks survey respondents the following dichotomous question: "Generally speaking, do you believe that most people can be trusted or you can't be too careful in dealing with people." Figure 1 shows the mean trust game allocations by country.[3] Additionally, we employ one of the largest datasets of aggregate trust game data behaviour from 167 trust game studies conducted in 36 countries.[4] The samples include developed and developing countries, studies in Africa, Europe, North America,

[3] For comparison purposes we rescaled allocations in the trust games to a 0-1 scale to match the range of the GSS question.

[4] The aggregate data is an updated and expanded version of Johnson and Mislin's (2011) dataset.

South America, and Asia, and include a variety of samples from student populations to nationally-represented samples.

Figure 1. Behavioural and survey measures of trust



2. Inter- and intra-country comparisons

10. As a starting point for understanding issues of measurement validity in a cross-national setting it is valuable to examine the convergence of survey and behavioural measures. To that end, we test the association between the GSS interpersonal trust question and Player 1 allocations from anonymous trust games in three ways: (1) in an individual-level cross-section; (2) at the aggregate level across cases (both within and out-of-sample); and (3) within individuals over time. By connecting these two measures we can start to see whether and where classic measures assumed to capture trust attitudes are, in fact, linked with what is assumed to be trusting behaviour in an economic, game-theoretic setting. We address these analyses in turn. We also examine the cross-country correlation between survey and experimental measures of interpersonal trust and GDP per capita. This provides a comparative test of the two measurements approaches from the perspective of construct validity. In addition, much of the interest in levels of trust derives from the belief that trust is an important driver of economic (and social) outcomes, so it is of particular interest to gain an understanding of the relative performance of the two measurement approaches.

11. Table 1 below reports correlations between the GSS and trust game measures from the pooled dataset containing national and student samples in seven countries

(ranging from developing democracies to OECD countries).[56] Overall we find a significant individual-level correlation between answers to the classic GSS question and trust-game behaviour. Individuals who expressed greater trust in the survey instrument also displayed greater trust in the trust game. Substantively, however, the association is modest (point-biserial $r = 0.11$) and, indeed, if we look country by country the strength of the association varies from $r = 0.07$ (and insignificant, South Africa) to $r = .30$ (Mexico). In general, then, the link between the GSS question and trust game allocations does not appear particularly strong at the individual level. This is in line with previous findings (Glaeser, 2000; Lazzarini et al., 2004).

12. This weak association between survey and game-based trust measures at the individual-level has hampered adoption of interpersonal trust measures by both researchers and statistical offices. However, this modest relationship may be substantially stronger at levels of higher aggregation (e.g. regional or national). Trust is typically defined as an expectation of positive behaviour on the part of others (Fehr, 2009; OECD, 2017), and generalised trust in others is fundamentally about a person's expectations with respect to society as a whole, i.e. "is the representative person in my society trustworthy?" Because of this, we would expect most meaningful variation in trust to exist between countries – where individuals will have reasonably different expectations of the trustworthiness of the representative person – rather than within countries. Studies examining the correlation between survey and experimental trust within a single experiment (or even different experiments within the same country) face the inherent limitation that all experimental subjects are interacting with a panel of people with the same *ex ante* level of expected trustworthiness. Correlation (or the lack thereof) between survey and experimental trust will depend entirely on the personality of the participant and cannot, by construction, reflect any variation in the expected trustworthiness of the person with whom the participant is interacting (in non-repeated anonymous games). If this is the case, then within-country studies of individual levels of interpersonal trust would not necessarily be expected to find significant correlations between methods.

13. To probe this possibility, Table 1 reports the correlation between sample aggregates of average trust (means) as measured by the GSS question and anonymous trust-game play in the seven countries (eight studies) with individual-level data. It is much larger ($r = 0.31$) than the one reported above but due to the small sample size ($n = 7$) it is statistically insignificant.

[5] Student samples: Chile 2009, US 2011, El Salvador 2010, Mexico 2012, South Africa 2011, Uruguay 2010. National samples: Spain 2012 (all waves), Portugal 2012 (does not include the GSS question). Adult, non-national sample: US 2015.

[6] The Spain and Portugal surveys were directed by Mariano Torcal with the financial support of The Spanish Ministry of Innovation and Science, research project references: CSO2009-14434 and CSO2013-47071-R.

Table 1. Test of association: Behavioural and survey measures of trust

Individual-Level Correlation			**National-Level Correlation**	
	Pearson's *r* and *p* value	*n*	Pearson's *r* and *p* value	*n*
Overall	.11	2375	.31	7
	.00		*.45*	
Chile	.15	287		
	.01			
El Salvador	.18	147		
	.03			
Mexico	.30	163		
	.00			
Portugal	n/a	473		
	n/a			
South Africa	.07	308		
	.23			
Spain	.05	620		
	.19			
US 2011	.05	158		
	.55			
US 2015	.10	461		
	.03			
Uruguay	.17	231		
	.01			

Note: Significant correlations are highlighted in yellow.

14. To assess convergence another way, we supplement the individual-level survey data with aggregate trust measures from 159 anonymous trust game studies conducted in 32 countries (to make 36 countries in total). That is, we match the mean player 1 allocation from each study with national mean scores on survey trust questions from the WVS and, where missing, the European Values Survey. Pair-wise correlation results in Table 2 show a moderately strong and significant relationship between aggregate measures of survey (classic GSS question) and game-based trust. The pooled correlation between game play and mean WVS values in the sample is $r = 0.22$.

15. Since some countries are significantly over-represented in the sample (for example, there are many more studies from the US in the sample than from any other country) this result may be biased by a few countries where much of the trust game research has occurred. To account for this we take the mean Player 1 allocation by country (pooling studies by country) and correlate it with the mean WVS value by country. By switching the level of analysis from studies within countries to country averages the correlation strengthens slightly to $r = 0.29$ from $r = 0.22$. This provides evidence that at the aggregate-level the mean trust measured in games is significantly

related to mean scores from the GSS/WVS question, albeit only at a moderate level. This result emerges despite the fact that most, but not all, of the trust games studies included were not nationally representative surveys (unlike the GSS/WVS data). Additionally, we included corrections (not shown) for sample type (student vs. non-student), payment type (lottery or not), whether real stakes are used, and whether subjects also played the role of Player 2 in the game. The inclusion or exclusion of these corrections do not substantively affect the results – national average scores on the GSS/WVS question correlate positively and reliably with mean trust allocation in anonymous trust games, regardless of the subject sample or protocol used in the games.

Table 2. Aggregate-level correlations: Behavioural and survey measures of trust

	Pearson's r and *p* value	*n*
GSS Question without Controls	.22	167
	.00	
GSS Question with Controls	.20	167
	.01	
GSS Question: country means	.29	36
	.08	
Trust: Family	.05	144
	.55	
Trust: Neighbourhood	.05	144
	.58	
Trust: Meet Personally	.17	144
	.04	
Trust: Meet First Time	.12	144
	.14	
Trust: Another Religion	.12	144
	.14	
Trust: Another Nationality	.10	144
	.22	

Note: Significant correlations are highlighted in yellow.

16. Next we examined if other formulations or scales of interpersonal trust questions better predicted trust game play than the classic WVS instrument. The results from both the aggregate data and from individual-level data from Spain show that other common survey questions are no better than the single WVS question.

17. One wave of the Spanish panel survey included the 3-question set of interpersonal trust questions used in the European Social Survey (among others) using 10-point scales.[7] Using either an additive scale composed of the three questions (α = .8; Principal Components Analysis suggests all variables load on a single dimension) or using the questions individually we find similar degrees of correlation with trust-game behaviour (r = 0.06) as with the GSS question and the trust game (r = 0.07). Furthermore, the ESS additive scale and the GSS trust question correlate at r = 0.53. This suggests that the

[7] This battery has also been used in the some waves of the WVS and versions of the GSS and Pew Social Trends survey: "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?"

multi-question battery may be capturing additional or different information than the single-question instrument, but this information is not related to behavioural trust captured in game play.

18. The evidence available from cross-country correlations between survey and experimental measures of trust outlined above is interesting but ambiguous. While a correlation between 0.22 and 0.29 is significant, it is not high in absolute terms. This leaves unanswered the question of what to make of the relatively weak evidence of convergent validity between survey and experimental measures. One possible approach, which we explore, is to compare the performance of the two measurement approaches with respect to construct validity (how well each measure conforms to the expected relationships with other variables).

19. Table 3 below presents the correlation between survey and experimental measures of trust and the natural logarithm of GDP per capita. Column I of Table 3 shows the results of an OLS regression with experimental trust measures (Player 1 allocations) as the independent variable, while column II repeats the same regression but with the proportion of the population indicating that people can be trusted in the standard survey question as the independent variable. Columns III and IV include both trust measures, with column IV excluding data from the United States.

Table 3. Trust measures as predictors of log GDP/capita

	I*	II*	III*	IV**
Game Allocations	.025		.012	.011
	.008		*.008*	*.011*
WVS Question		.042	.042	.038
		.009	*.009*	*.009*
R^2	.08	.18	.24	.21
N	163	163	163	117

Note: Significant coefficients are highlighted (p<.05). OLS regressions with robust Standard Errors in Italics. *Regressions include controls for receiver endowment and random payment. **Regression includes controls and excludes the United States from the sample.

20. It can be seen from Table 3 that both survey and experimental measures of trust have a significant correlation with GDP per capita in regressions I and II. Of the two measurement approaches, however, survey measures of trust perform significantly better, explaining roughly twice the proportion of variance in GDP (R^2 = 0.18) compared to experimental measures (R^2 = 0.08). When both survey and experimental measures of trust are included in the regression at the same time (regressions III and IV), however, the experimental results cease to be significant. This holds whether data from the United States is included in the regressions or not.[8]

21. Taken at face value, Table 3 would seem to indicate that survey measures of trust out-perform experimental measures of trust in terms of their correlation with GDP per capita. However, a note of caution is warranted. Many of the studies included here are

[8] Because many of the experimental studies included in the dataset used here originate in the United States, it may be that data from the United States are biasing the regression. Re-running the data omitting studies from the United States addresses this issue.

based on data from student samples that may not be representative of the population as a whole. In particular, there is real risk that samples of tertiary students may share important values or norms with tertiary students in other countries compared to the bulk of the population in their own country. Table 4 below examines this by repeating the regressions in Table 3 but omitting all those studies using student samples. This decreases the sample size by approximately 75%.

Table 4. Trust measures as predictors of log GDP/capita, excluding student samples

	I*	II*	III*	IV**
Game Allocations	.075		.058	.064
	.013		*.02*	*.018*
WVS Question		.041	.025	.02
		.012	*.014*	*.012*
R^2	.26	.20	.32	.35
N	40	40	40	37

Note: Significant coefficients are highlighted (p<.05). OLS regressions with robust Standard Errors in Italics. *Regressions include controls for receiver endowment and random payment **Regression includes controls and excludes the United States from the sample.

22. In general, the patterns in Table 4 repeat those in Table 3. Both the experimental measures and survey measures are significantly correlated with log GDP per capita in regressions I and II, while only one of the measures is significant in columns III and IV. The main difference between Tables 3 and 4 is that with student samples omitted, it is the experimental results that have a stronger correlation with log GDP per capita in columns I and II, and it is the experimental measure that remains significant in columns III and IV.

23. The available data clearly show that both experimental and survey measures of trust exhibit the expected relationship with GDP per capita. However, no clear picture emerges as to which measurement approach performs best overall, with results being sensitive to the inclusion of student samples. One important finding, though, is consistent regardless of the inclusion of student samples: under no regression are both survey and experimental trust significant at the same time. While this result should not be over-interpreted, it is consistent with the view that both survey and experimental measures of trust are associated with higher levels of GDP per capita via the same mechanism.

3. Measurement reliability of trust measures

24. The finding of a moderate aggregate-level correlation between survey and game-based measures but only a weak individual-level one raises the question: why the disconnect? One likely explanation is measurement reliability or noise. Fortunately, we can assess the reliability of the measures since our data includes national-sample panel studies involving multiple waves of trust games with the GSS question or the GSS alone. Overall we find moderate test-retest measurement reliability for both measures.[9] In a Spanish panel study fielded in 2012 the GSS question registers an overtime interclass correlation of $r = 0.46$ over a period of seven months. For the anonymous trust game, the two-wave interclass correlation is $r = 0.41$. We find similar evidence of the test-retest reliability of the GSS question from a panel study in Colombia. In that sample, the two-wave interclass correlation for the GSS question is $r = 0.42$ over a period of seven months. In both cases the trust games or GSS questions were anonymous and included no social identity cues. Overall we find only moderate and similar degrees of test-retest reliability for both measures in two different contexts. Of course, these moderate reliability scores may also reflect reality: respondents' interpersonal trust levels may have actually shifted between survey waves. However, the theoretical and empirical literature on interpersonal trust consistently argues that interpersonal trust is stable trait that is unlikely to shift substantially over a period of weeks or months.

25. These results fit with evidence from the 1972-1974-1976 American National Election Study (ANES) panel survey. Despite being interviewed years apart, respondents answers to the GSS question correlated with prior answers at a degree similar ($r = 0.41$) to those found in the short panel times used in Spain and Colombia. Uslaner (2002) makes the assumption that the GSS trust question involves no error and that the correlation indicates that the attitude exhibits a good deal of overt-time stability. But, as our data from Spain and Colombia indicate, the stability of survey measures is only moderate, and changes in measured survey trust are likely to involve significant level of noise. Since the short panels used in Spain and Colombia are much more likely to capture random changes in question response than true attitudinal changes, individuals' attitudes regarding trust could even more stable than the ANES panel data indicates.

26. Not only are the Spain and Colombia studies well suited to measuring instrument reliability, they can also be used to estimate – and correct for – measurement error. While linear model estimates allow for random error in the outcome measure (game transfers in our case) they assume no measurement error in the explanatory variables (GSS question). If such random noise exists in the independent variable it truncates the observed relationship and often biases estimates downwards, known as attenuation bias. Because we have an estimate of random measurement error for the GSS question we can use an errors-in-variables regression (or measurement error model) approach to account for attenuation bias caused by random measurement error.[10] That is, using the average of the

[9] Test-retest reliability is the standard technique for measuring the reliability of survey questions (see, for example, Krueger and Schkade, 2008). Alternative approaches, such as looking at the ratio of the variance of one variable to the covariance of the measures produce results with similar conclusions (survey trust: 0.21, trust game allocation: 0.23).

[10] In undertaking such an analysis, we do not include the information from the ANES panel because the long time between waves means that it is more difficult to attribute changes in

interclass-correlation estimates from the Spain and Colombia studies (r = 0.44) as a measure of reliability for the GSS question allows us to estimate a corrected (i.e. adjusted for measurement error) relationship between the survey- and game-based trust measures. Specifically, in an errors-in-variables approach the average test-retest correlations for the GSS/WVS question in Spain and Colombia are used as an estimate of measurement reliability.

27. Since the coefficient for a variable in bivariate linear regression is in part a function of the variance of the independent variable (i.e. $\beta = \frac{cov(x,y)}{var(x)}$), we can use our estimate of measurement reliability to parse the true variance of the GSS question from the random measurement noise[11] at the individual-level. Using this errors-in-variables approach causes the correlation between the GSS question and anonymous game allocation to nearly double – the pooled estimate of correlation between the survey and games measures in the eight studies with individual-level data rises to r = 0.17, which is close to the results from the aggregate data analysis presented above (r = 0.22). Table 3 displays the results. This provides further evidence, from the individual-level, that survey trust questions, in particular the GSS question, significantly and consistently measure, at least in part, the same underlying concept of trust as game play.

Table 5. Test-retest reliability and errors-in-variables results

	Interclass-correlation and *p* value	*n*
GSS: Colombia	.42	1023
	.00	
GSS: Spain	.46	228
	.00	
Game: Spain	.48	228
	.00	

	Pearson's *r* and *p* value	*n*
Errors-in-Variables Model	.17	2375
	.00	

Note: Significant correlations are highlighted in yellow.

responses to measurement error or valid attitude changes. Another way to conceive of measurement error models is as a latent variable (true X) variable approach with reliability estimates functioning as an instrument.

[11] Another way to conceive of measurement error models is as a latent variable (true X) variable approach with reliability estimates functioning as an instrument.

28. In a related way, we can test how game play and survey questions are linked by priming the trustor (Player 1) with information about the social identification of the trustees (Player 2). By priming the identification of whom the subjects are asked to trust (both in the game and survey question) we expect to decrease response variability, similar to the econometric results from the Errors-in-Variables approach above (cf. Zaller and Feldman, 1992).

29. One wave of the Spanish survey included trust games *and* GSS-style survey questions in which trustors are given identifying characteristics of the trustees. Specifically, they were informed about regional identity (e.g. Catalan, Basque, Madrileños) and partisan identity (e.g. PP, PSOE, etc.), both of which are found to condition trust (Carlin and Love, forthcoming; Martini and Torcal, forthcoming). That is, appropriating the GSS framework, the survey questions asked subjects if, say, a Madrileño can be trusted or you can't be too careful. In the trust games, social identity was primed before the trust decision was made. In contrast to the anonymous game/question, the primed treatments help ensure that all subjects hold in their mind a similar group when making the trust decision. Since the anonymous game allows individuals to imagine a range of possible trustees the primed games/questions reduce it to one type of possible trustees, thus reducing response variance in comparison to the anonymous game. Results indicate that this prime significantly strengthens the link between question and game results. Individual-level correlation between identified GSS questions and identified trust game behaviour is, on average, $r = 0.40$, dramatically higher than the individual-level correlation between the anonymous GSS question and the anonymous game in Spain ($r = 0.05$).

4. Implications for trust measures

30. Information on interpersonal trust is of potentially great interest to policy-makers, economists, and other academics concerned with the drivers of key social and economic outcomes. The available trust data are highly suggestive and provide good prima facie evidence that intangible factors such as trust may account for much of the variation in social and economic outcomes that cannot be accounted for through traditional socio-economic variables. However, the value of this analysis depends crucially on the validity of the trust data underpinning it.

31. The lack of consistent findings on the correlation between survey measures of trust and implementations of the trust game in the literature raises some reasonable doubt over the convergent validity of existing trust measures. This, in turn, acts as a significant disincentive for the development of better data sets – particularly from national statistical offices. Yet it is precisely this sort of data that is needed to inform policy, methodology, and theory. There is thus a real risk of a vicious cycle whereby methodological uncertainty prevents the collection of the data needed to resolve the main methodological challenges. This paper contributes to resolving this *impasse* by drawing together a large number of experimental studies in a single dataset that allows for the analysis of the relationship between survey and experimental measures of trust both at the individual *and*

aggregate country levels. Because the dataset allows analysis at both levels it is possible to investigate whether the differences between authors using individual level within-country data such as Glaeser (2000) (no correlation between survey and experimental data) and authors using aggregate between-country data such as Johnson and Mislin, (2011 and 2012) (robust correlation between survey and experimental data) are a function of the different level of analysis or reflect more fundamental measurement problems.

32. Examining survey and experimental data at the individual and aggregate levels identifies two main findings.

- First, the data confirm the results of both Glaeser et al. *and* Johnson and Mislin. The standard survey question regarding generalised trust in others and experimental behaviour revealed through the trust-game are at best weakly correlated at the individual level. However, they show a consistent and significant, although still only moderate, correlation at the cross-country level. At the individual level, this correlation strengthens dramatically once measurement error common in surveys are accounted for. This is consistent with the view that much of the meaningful variation in generalised trust varies at the level of society and that individual within-country variation in generalised trust carries less information. An analysis of the relationship between survey and experimental measures of trust and GDP per capita across countries supports this conclusion. If this is the case, the diverging results of Glaeser et al. and Johnson and Mislin should not be considered as indicating a fundamental problem with survey measures of trust, but instead as reflecting the nature of the phenomenon being measured.
- A second important finding relates to measurement issues. The standard Rosenberg question on generalised trust is often criticised for being poorly worded, unclear, and difficult to translate. This has led to a number of attempts to develop better questions. However, analysis of the dataset used in this paper highlights that newer survey questions designed to replace or supplement the Rosenberg question, such as the ESS/WVS 3 question battery, generally do not more strongly converge with trust-game behaviour than the original question formulation.

33. The analysis also indicates the value of large, comparable, cross-country datasets on interpersonal trust and the limitations of more narrowly focused small sample datasets. Unfortunately, the vast majority of experimental measures of trust come from small sample within-country studies; even the datasets such as the one used here, which draws together experimental data on trust from a wide range of small sample datasets have obvious limitations. Further improvements in the measurement of trust therefore depend on progress in two areas. The first is better quality survey data – ideally through datasets of comparable scale and quality to those used by national statistical offices. Second, it is crucial to acquire large, comparable, nationally representative datasets of experimental results that can be used to validate survey data without the limitations associated with the use of small convenience samples such as those lying behind the dataset used here.

References

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

Berg, J., J. Dickhaut and K. McCabe (1995), "Trust, reciprocity, and social history", *Games and Economic Behaviour*, No. 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 Paper*, No. 2012/03, OECD Publishing, Paris, http://dx.doi.org/10.1787/5k9b9ltjm937-en.

Boreham, R. et al. (2002), "Health Survey for England 2000: Social capital and health", in G. Prior and P. Primatesta (eds.), *Health Survey for England 2000,* The Stationery Office: London

Carlin, R.E. and G.J. Love (forthcoming), "Political competition, partisanship, and interpersonal trust in electoral democracies: Partisan trust bias in comparative perspective", *British Journal of Political Science.*

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

Ginn, J. and S. Arber (2004), "Gender and the relationship between social capital and health", in 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.

Haas, B.W. et al. (2015), "The tendency to trust is reflected in human brain structure", *NeuroImage*, No. 107, pp. 175-181.

Harris, L.T. et al. (2007), "Regions of the MPFC differentially tuned to social and nonsocial affective evaluation", *Cognitive, Affective, and Behavioural Neuroscience*, Vol. 7, pp. 309-316.

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

Johnson, N. and A. Mislin (2012), "How much should we trust the WVS trust question?", *Economics Letters*, No. 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.

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

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

Krueger, A.B. and D.A. Schkade (2008), "The reliability of subjective well-being measures", *Journal of Public Economics*,Vol. 92, No. 8, pp. 1833-1845.

Krueger, F. et al. (2007), "Neural correlates of trust", *Proceedings of the National Academy of Sciences* Vol. 104, No. 50, pp. 20084-20089.

Lazzarini, S. et al. (2004), "Measuring trust: An experiment in Brazil", *Insper Working Paper*, No. 049/2004.

Martini, S. and M. Torcal (forthcoming), "Trust across political conflicts: Evidence from a survey experiment in divided societies", *Party Politics*.

OECD (2017), *OECD Guidelines on Measuring Trust*, OECD Publishing, Paris, (forthcoming).

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

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

Sampson, R.J. (2012), *Great American City: Chicago and the Enduring Neighborhood Effect*, www.press.uchicago.edu/ucp/books/book/chicago/G/bo5514383.html.

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

Stafford, M. et al. (2004), "Neighbourhood social cohesion and health: Investigating associations and possible mechanisms", in A. Morgan & C. Swann (Eds.), So*cial Capital for Health: Issues of Definition, Measurement and Links to Health*, pp. 11-132, London: Health Development Agency.

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

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

Zaller, J. and S. Feldman (1992), "A simple theory of the survey response: Answering questions versus revealing preferences", *American Journal of Political Science*, Vol. 36, No. 3, pp. 579-616, www.uvm.edu/~dguber/POLS234/articles/zaller_feldman.pdf.