Article

Validating a scale for citizen trust in government organizations

Stephan Grimmelikhuijsen and Eva Knies

Utrecht University School of Governance, The Netherlands

Abstract

Citizen trust in government at the macro level has been studied by public administration scholars for many years. To further our understanding, assessing trust at the meso level of government organizations is important to more precisely determine effects and antecedents of trust at the organizational level. The organizational trust literature has shown that organizational trustworthiness is multidimensional, but the extant literature has not validated such measures in a public administration context. The proposed scale builds on and adapts an existing organizational trust scale to a public administration context. The 'Citizen Trust in Government Organizations' scale is validated using data from two different samples (total $n = 991$), resulting in a scale of nine items measuring three dimensions: perceived competence, benevolence, and integrity. This scale can be used by other researchers and is valuable to gain a more specific and multi-dimensional understanding of trust in government organizations.

Points for practitioners

A major problem for government organizations worldwide is the lack of perceived trustworthiness by the public. To tackle this problem, a way to measure it is needed, but at the moment there are only generic measures to assert perceived trustworthiness in a government organization. This article presents a first validation and incorporates three dimensions: perceived competence, benevolence, and honesty. Practitioners can use this scale and adapt to their relevant local context to identify specific trustworthiness problems.

Corresponding author:
Dr Stephan Grimmelikhuijsen, Utrecht University School of Governance, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands.
Email: s.g.grimmelikhuijsen@uu.nl

Introduction

Citizen trust in government has been an important topic of study for public administration and political science scholars for many years (e.g. Carnevale, 1995; Dalton, 2004; Fukuyama, 1995; Hetherington, 1998; Levi and Stoker, 2000; Miller, 1974; Nyhan, 2000; Putnam, 2000; Tolbert and Mossberger, 2006; Van der Meer, 2010; Vigoda-Gadot et al., 2010; Yang and Holzer, 2006). This has led to important publications on, for instance, the determinants of trust in government (Bovens and Wille, 2008; Hibbing and Theiss-Morse, 2001; Norris, 1998; Nye et al., 1997; Van de Walle et al., 2008; Vigoda-Gadot, 2007; Vigoda-Gadot et al., 2007), and on the effects of trust on society (e.g. Fukuyama, 1995; Putnam, 2000).

It is common practice to measure trust in government *in general* using single-item measures of political trust (Hooghe, 2011: 275; Van de Walle, 2004) or the one-dimensional multiple-item measure of the American National Election Studies (Citrin, 1974; Poznyak et al., 2014). In contrast, studies measuring trust in a particular organization of government (e.g. *this* municipality or *this* public service) are scarce. Van de Walle (2004: 146) noted that it is important to trace specific objects of trust to determine the causes of trust in government organization more precisely, as the antecedents of generalized trust and trust in a specific organization are not necessarily the same (see also Van de Walle et al., 2005).

In addition, Kim (2005: 614) notes that public administration scholars have not yet found an appropriate model of public trust. Assessing trust at the meso level, i.e. trust in a specific government organization, may help in addressing this lacuna: measuring trust in government organizations better allows us to identify particular organizational contexts that relate to the antecedents of trust in that organization. As a consequence, understanding trust in specific organizations can help to develop more appropriate models for public trust in general.

However, if we want to measure trust in a specific government organization we need a different measure than the one-dimensional measure that is common in current research on trust at the macro level. The extant literature on organizational trust has provided extensive evidence that organizational trust is based upon the perceived trustworthiness of an organization and that this consists of multiple dimensions. Given the complexity of the construct, it indeed seems most appropriate to assess trustworthiness empirically (McEvily and Tortoriello, 2011; Mayer et al., 1995).

The multi-dimensional nature of organizational trustworthiness has been validated empirically in various studies (e.g. Cummings and Bromily, 1996; McKnight et al., 2002; Mayer and Davis 1999), but not in a public administration context. A validation for this context is needed, because trust researchers have acknowledged that trust is inherently context-dependent (Bachmann, 2011; Granovetter, 1985; Hardin, 2002; McEvily and Tortoriello, 2011; Rousseau et al., 1998). According to Bachmann (2011: 211), institutional context 'shapes the way that actors make their decisions, including the decision to either trust or not to trust another actor'. In theory, taking context into account could lead to unique

measures for each specific organization. On the other hand, there is a need to build a cumulative body of knowledge by using existing and validated measures (McEvily and Tortoriello, 2011). To address a balance between these two ends, we use an existing validated scale – to contribute to knowledge accumulation, but adapt it to a public administration context – to address the contextual nature of trust.

In this article we focus on measurement of trustworthiness, as it is a crucial basis for trust in government organizations (Mayer et al., 1995). In line with insights from organizational trust research we conceptualize this as a composite of three dimensions: perceived competence, perceived benevolence and perceived integrity (e.g. McEvily and Tortoriello, 2011; McKnight et al., 2002; Mayer et al., 1995). These three dimensions of perceived trustworthiness are key elements that form the basis for other facets of trust, such as trusting intentions and behavior (McEvily and Tortoriello, 2011). Based on these theoretical notions we apply a psychometric scale validation procedure using two different samples ($n = 991$), which results in a multi-dimensional measurement scale of nine items that can be used to gain a more specific understanding of organizational trust in the context of public administration. We specifically test the scale in municipal government as a crucial case for validation, as municipalities have the ties closest to citizens, and citizens have a more direct stake in local issues (Pina et al., 2007; Piotrowski and Van Ryzin, 2007). Although our scale test is carried out in a municipal context, our scale uses templates to leave room for flexible measurement of various public sector organizations.

Conceptualizing trust

Three important and internationally widely used trust in government measures are the Eurobarometer for European comparisons, World Values Survey for worldwide comparisons (e.g. Van de Walle et al., 2008), or Pew and ANES for US-based analyses (see, for critical discussions on ANES: Citrin, 1974; Poznyak et al., 2014). Many researchers use these indicators to measure trust in government in general. However 'trust in government' in general refers to an amorphous entity and gives a very generic understanding of its effects and antecedents (Fisher et al., 2010). Especially when a single item measure is used, the ambiguity of the concept may mean different things to different people.

To gain a more specific understanding of how trust works, researchers need to focus on trust in specific government services or organizations (cf. Van de Walle, 2004; Van de Walle et al., 2005). However, macro-level single-item measures are too limited to measure trust in specific organizations, as a number of analyses on organizational trust have established evidence that organizational trust depends on perceived trustworthiness, which consists of multiple dimensions (e.g. McAllister, 1995; McEvily and Tortoriello, 2011; Mayer et al., 1995).[1]

There is no full agreement on how to conceptualize trust versus trustworthiness exactly (Bachmann, 2011; McEvily and Tortoriello, 2011), but we view trust as a

characteristic of a trustor (I have trust in ...), whereas perceived trustworthiness emphasizes the perceived characteristics of the trustee (I think X is trustworthy). These characteristics are important to understand why some organizations are more trusted than others. Although perceived trustworthiness is not trust per se, it helps to build the foundation to further develop trust (Mayer et al., 1995: 717).

To understand the multidimensionality of perceived trustworthiness we will first provide a definition of trust in general. Given the complexity of the concept, and the multiple disciplines that have embarked on trust research, defining trust is not easy. Understanding why and how people trust others has been a central focus of research for psychologists, sociologists, political scientists, economists, and organizational scientists. Across disciplines and even within disciplines, a myriad of definitions, concepts, and operationalization are being used in research.

That said, there seems to be cross-disciplinary agreement about two elements that are relevant to trust: a degree of 'risk' and 'interdependence' (Bachmann, 2011; Fisher et al., 2010; Hardin, 1993; Lewicki and Bunker 1995; Lewicki et al., 1998; Luhmann, 1979; McAllister, 1995; Mayer et al., 1995; Zucker, 1986). According to Hardin (1993), trust is a threefold relationship, in which A trusts B to do X. A expects B to do a certain thing which is in his/her interest. In case of citizen trust in government, this threefold relation yields a risk as citizens are uncertain as to whether government organizations actually carry out the tasks they are entrusted to. Risk becomes relevant when governments exert a certain degree of power over citizens, which can either be used properly or abused.

The condition of interdependence implies that the interests of one party cannot be achieved without reliance on the other party. So, in the case of trust in government, if citizens want government to solve pressing social problems, they are often dependent on government organizations to deliberate on decisions, carry out policy measures, and monitor their effects. Government, on the other hand, depends on citizens to cooperate and act according to certain rules for its policies to have any effect (Ayres and Braithwaite, 1992; Levi and Stoker, 2000).

If these two conditions are present, then trust becomes relevant in a relationship. Based on these conditions, Mayer et al. (1995: 712) developed a much-cited definition of trust: trust is 'the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party'.

This definition is widely used, and serves the purpose of this study very well. It was explicitly meant to be generalizable to various levels of analysis. Therefore, it is a definition that can be used for studies at various levels of analysis, such as a group trusting a group, an individual trusting an individual, or – such as in our case – an individual trusting an organization.

In this definition, the expectation of the vulnerable party (i.e. citizen) is central. These expectations are based on the perceptions that people have of the 'other'; in other words: are the intentions and behaviors of the other perceived to be trustworthy? (see Hardin, 2002). This implies that 'trust in a government organization'

actually consists of the extent to which it is considered 'worthy of trust' by its citizens. Perceived trustworthiness is not seen as an objective quality of government, but as an aggregate of perceptions of citizens (Bouckaert and Van de Walle, 2003).[2] According to the literature on organizational trust, we best conceptualize perceived trustworthiness in multiple dimensions (McEvily and Tortoriello, 2011). Various literature reviews on organizational trust have shown that the dimensions of perceived competence/ability, benevolence, and integrity are recognized to be central dimensions in most organizational trust studies (McEvily and Tortoriello, 2011; McKnight et al., 2002; Mayer et al., 1995). The most recent review by McEvily and Tortoriello (2011) indeed shows that the top three dimensions are integrity (19 times), ability/competence (14 times), and benevolence (14 times).

These three dimensions can be defined as follows for government organizations:

- *perceived competence*: the extent to which a citizen perceives a government organization to be capable, effective, skillful, and professional;
- *perceived benevolence*: the extent to which a citizen perceives a government organization to care about the welfare of the public and to be motivated to act in the public interest; and
- *perceived integrity*: the extent to which a citizen perceives a government organization to be sincere, to tell the truth, and to fulfill its promises.

These three dimensions have formed the basis for various noteworthy measurement scales (Cummings and Bromily, 1996; McAllister, 1995; McEvily and Tortoriello, 2011; Mayer and Davis, 1999; Mayer et al., 1995). However, these measures focus either on interpersonal managerial trust (McAllister, 1995) or intra-organizational trust (Cummings and Bromily, 1996; Mayer and Davis 1999), whereas our purpose was to test a measure for an external actor (i.e. citizens) in a government organization. An important scale for measuring trust of external actors in an organization has been developed by McKnight et al. (2002). This is a much-used measurement scale in marketing and e-commerce (cited 2185 times in Google Scholar and 614 times in Web of Science). Based on the above-mentioned dimensions of trustworthiness, McKnight et al. (2002) validated a scale measuring perceived trustworthiness of customers in commercial organizations.

The empirical context in which a commercial enterprise operates is, however, different to that of a government organization (Perry and Rainey, 1988). Firstly, customers have more freedom of choice and as such are less dependent on one particular organization. Second, the role of a commercial enterprise in the various dimensions of organizational trust is supposedly different. Especially with regard to benevolence, one would expect a government organization to act in the interest of citizens in general (i.e. public interest), whereas a commercial enterprise is focused on a narrower interest – of individual consumers (i.e. private interest). These crucial differences should be taken into account when measuring trust in a government organization.

Whether this contextual difference matters for measuring trust is subject to debate. Scholars from organizational psychology often argue that trust is a psychological trait and as such it can be conceptualized relatively free of context (Bachmann, 2011: 205). In those instances, the institutional context is seen as something that is external to the individual psychological make-up. On the other hand, sociologists find that the institutional context is interwoven with trust and is actually something that generates trust (Bachmann, 2011; Granovetter, 1985). McEvily and Tortoriello (2011: 41) propose a middle course and argue that the empirical setting and context are indeed relevant for measuring trust but that trust is not completely idiosyncratic to each specific setting: 'where commonalities exist, previously validated measures should be adopted rather than developing a new measure'.

We concur with this middle course perspective on the role of context in measuring trust. Therefore, our scale is based on an existing organizational trust scale by McKnight et al. (2002) to contribute to the existing body of knowledge on organizational trust. We chose this scale because it is one of the few validated scales that measure individual trust in an organization and it is widely used in other fields of research. However, the scale is adapted for research in a public administration context to do justice to the contextual nature of trust. In doing so we balance the contextual nature of trust and the need to build on existing, validated measures. Before we present this measure, the next section outlines the process of psychometric scale validation. This procedure is a powerful tool to determine whether a measurement scale is indeed reliable and valid.

Method

The empirical adaptation and validation of the trust scale involved several steps. We followed the often-used procedure outlined by DeVellis (2003). The data analysis was carried out using Mplus (Muthén and Muthén, 2012).

Phase 1: Item generation

The goal of the first phase in the validation procedure was to generate items measuring the three trust dimensions. In phase 1, we used the McKnight et al. (2002) items and adapted them. Our strategy was to not change the items unless it was necessary for them to fit a public context. This way we tried to strike a balance between closely adhering to the current body of research on organizational trust and, at the same time, paying attention to the contextual nature of trust. A number of experts on trust in government and on research methodology reviewed it for content and face validity.

Phase 2: Psychometric properties of the scale

The goal of the second phase was to test whether the generated items indeed measure the hypothesized dimensions of trust and to determine the reliability of

the scale. The psychometric testing in phase 2 involved several steps. First, we examined the dimensionality by performing an exploratory factor analysis (EFA). Second, we conducted confirmatory factor analyses (CFAs) in order to examine the fit of our three-dimensional model and to compare it to several other one- and two-factor models to test its robustness. Third, we assessed reliability by examining the coefficients of Cronbach's alphas.

Phase 3: Convergent validity of the scale

The goal of the third phase was to determine whether our multidimensional measure of trustworthiness is significantly related to prior measures of trust in government (convergent validity). If there was no relationship between our multidimensional scale and previously used measures of trustworthiness, one could question whether we were measuring what we intended to measure. To determine convergent validity, we examined the relationship between the three dimensions of trust on the one hand, and on the other a one-item measure of trust ('I have trust in the XX, when it concerns XXX', measured on a 5-point scale), which is often used in public administration research. We expect a positive relationship between this item and our three dimensions.

Research sample

In this article, we used respondents from two samples. In total, 991 respondents participated in our study. The research samples are drawn from data gathered in the Netherlands, which were part of a PhD thesis. These data were collected in studies using an experimental design in which people were asked about their trust in local government in the Netherlands. The sample from study 1 consisted mostly of university students. The sample from study 2 consisted of a diverse group of citizens participating in a municipal citizen panel. The fact that our data come from two different samples allows us to replicate and cross-validate our findings.

All respondents filled out surveys including our trust scale (which will be presented in the next section), and a one-item trust measure often used in current research on trust for assessing the convergent validity of our scale.

Table 1 shows that both samples consist of respondents that vary on some relevant background characteristics. Approximately 45.0 percent of the respondents in both samples are male. The average age varies across samples as respondents in the first sample are over ten years younger than in sample 2 (mainly student population). Furthermore, although both samples draw heavily on highly educated respondents, sample 1 contains somewhat more highly educated respondents. The final relevant background trait shown in Table 1 is political preference, which differs between samples: in sample 1, 47.3 percent have a left-wing political preference, compared to 63.2 percent in sample 2. It should be noted that these are non-random samples, particularly sample 1. However, because this is

Table 1. Sample compositions

	% male	Av. age, years	% highly educated	Pol. pref. (% l.w.)
Sample 1 ($n = 156$)	44.9	29.7	86.5	47.3
Sample 2 ($n = 835$)	45.0	40.9	75.6	63.2

Pol. pref. (% l.w.) = Percentage of participants who indicated a preference to vote for a left-wing political party. 'Left-wing political parties' = D66, PvdA, GL, SP, PvdD (based on manifesto research by Klingemann et al., 2006).

a scale validation we are mostly interested in how the scale performs in various different subsamples.

Results of the scale validation procedure

In this section we present the results of the empirical validation of our three-dimensional trust scale. Successively, we will present the items that were generated (phase 1), the psychometric properties (phase 2), and the convergent validity of the scale (phase 3).

Phase 1: Item generation

The goal of the first phase in the validation procedure was to operationalize the three dimensions of perceived trustworthiness. Items were generated based on items by past research of McKnight et al. (2002) regarding trust in commercial organizations. This has been used because it is one of the few studies that contain a fully validated measurement of external trust in organizations. Because of the different context of private organizations, the items were adapted to the public context (see Table 2). Items were formatted as 5-point Likert scales: strongly disagree, disagree, neutral, agree, and strongly agree. This resulted in a 12-item questionnaire that was discussed with two academic experts on trust in government, and two methodological experts. Next, the questions were tested by a panel of ten testers, to assert whether all items were clear, which was indeed the case.

Table 2 shows 12 generated items to measure the three underlying dimensions: benevolence (3 items), competence (5 items), and integrity (4 items). As Table 2 shows, we tried to closely follow the original items. Most notable changes are made in the benevolence items. Our main idea behind departing from the original scale is that trust measurement should be adapted according to its relevant context (Bachmann, 2011; McEvily and Tortoriello, 2011). Looking at the specific context of public policy, our departure from the original operationalization of benevolence is justifiable because the government is not supposed to work in the best interests of a specific individual, but is defined and legitimated because collective goods or commons need to be governed (e.g. Ostrom, 1990). A key element why government is needed often lies within the need to organize or regulate such collective action.

Table 2. Items measuring the three dimensions

McKnight et al., 2002 items	Our items	In final scale
NO PREFIX	ADDED PREFIX: When it concerns [air quality policy] . . .	*
Overall, LegalAdvice.com is a capable and proficient Internet legal advice provider.	COMP1: [The municipality of XX] is capable.*	*
LegalAdvice.com is competent and effective in providing legal advice.	COMP2: [The municipality of XX] is effective.	
	COMP3: [The municipality of XX] is skilful.	
In general, LegalAdvice.com is very knowledgeable about the law.	COMP4: [The municipality of XX] is expert.*	*
LegalAdvice.com performs its role of giving legal advice very well.	COMP5: [The municipality of XX] carries out its duty very well.*	*
If required help, LegalAdvice.com would do its best to help me.	BEN1: If citizens need help, [the municipality of XX] will do its best to help them.*	*
I believe that LegalAdvice.com would act in my best interest.	BEN2: [The municipality of XX] acts in the interest of citizens.*	*
LegalAdvice.com is interested in my well-being, not just its own.	BEN3: [The municipality of XX] is genuinely interested in the well-being of citizens.*	*
LegalAdvice.com is truthful in its dealings with me.	INT1: [The municipality of XX] approaches citizens in a sincere way.*	*
LegalAdvice.com is sincere and genuine.	INT2: [The municipality of XX] is sincere.*	*
LegalAdvice.com would keep its commitments.	INT3: [The municipality of XX] keeps its commitments.	
I would characterize LegalAdvice.com as honest.	INT4: [The municipality of XX] is honest.*	*

Because of this crucial difference as compared to organizations in private contexts, we adjusted the benevolence scale.

Furthermore, we changed COMP2, by explicitly mentioning 'duty', which more closely fits with the public sector context in which organizations have certain duties that are required by law. Moreover, we rephrased items that mentioned two different terms in one item (i.e. LegalAdvice is competent and effective) so that the new items became unequivocal. The final important change made was the use of a 'prefix' which allows for specific domains and contexts to be questioned. In doing so we used templates. Templates allow for more flexibility and tailor-made

questions. Furthermore, it makes it easier for respondents to comprehend and relate to the items, and as such improves reliability and content validity (DeVellis, 2003: 52). An example of a template is as follows: [When it concerns domain X], [organization X] is effective.

The [first part] can be replaced with various domains. The [second part] specifies the name of the organization, or type of organization. For example, a template could be 'filled' as follows: When it regards environmental policy, the municipality of Amsterdam is effective.

Phase 2: Psychometric properties of the scale

The goal of the second phase was to test whether the generated items indeed measure the hypothesized dimensions of trust and to determine the reliability of the scale.

First, we examined the dimensionality of the trust items by performing an exploratory factor analysis, using Mplus (Muthén and Muthén, 2012). We used direct oblimin rotation because this allows for the factors being correlated, which is expected to be the case for the dimensions of trustworthiness. Typically, questionnaire items measured on a 5-point Likert scale are treated as continuous variables. However, our data were not normally distributed. In case of a non-normal distribution it is advisable to treat items in the analysis as categorical instead of continuous. Therefore, we decided to identify all our observed variables as categorical data.

To determine the model fit, we examined the comparative fit index (CFI), Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). Acceptable fit is evidenced by a CFI and TLI of .90 or higher (Bentler, 1990) and an RMSEA of .1 or lower (Browne and Cudeck, 1993). To determine the relative quality of our models and to provide a means for model selection, we used the Akaike information criterion (AIC) and Bayesian information criterion (BIC) fit indices.

In the exploratory factor analysis (EFA) we compared a one-factor, a two-factor, and a three-factor solution to assess if our assumption of a three-dimensional structure indeed fits with the data. The three-factor solution showed the best fit (CFI = .988; TLI = .976; RMSEA = .100; AIC = 19,468; BIC = 19,865) compared to the one-factor (CFI = .946; TLI = .935; RMSEA = .166; AIC = 20,147; BIC = 20,441) and two-factor solutions (CFI = .976; TLI = .964; RMSEA = .124; AIC = 19,688; BIC = 20,036). The factor structure was as we had expected, although we had to delete three items because these loaded on more than one dimension and/or had factor loadings below .5. We excluded two items from the competence dimension (COMP2 and COMP3) and one item from the integrity dimension (INT3).

This can also be explained conceptually. INT3 (honoring commitments) may be considered to be a 'hybrid' item, because honoring commitments may be related to being *unable* to honor it (related to incompetence) or deliberately dishonoring commitments toward citizens (related to a lack of integrity). The misfit of COMP2 (effectiveness) and COMP3 (skillfulness) is less easily explained. These

are items that depart somewhat from the original competence scale by McKnight et al. (2002), which may mean that they measure something different than perceived competence. For instance, most perceived competence items measure how well a government organization carries out its operations, whereas COMP2 asks about its effectiveness, which is more closely related to setting the right goals. Hence, perceived competence is a measure of 'doing things right', whereas COMP2 gauges whether government 'does the right things'. As regard to COMP3 (skillfulness) we do not have a clear-cut explanation. We speculate that it could be that skillfulness is more likely to be associated with trust in individuals, and less so to trust at the organizational level.

Second, we examined the dimensionality of the trust items again by performing a confirmatory factor analysis (CFA), using Mplus (Muthén and Muthén, 2012). Based on the results of our literature study and the EFA, we tested a model in which three items loaded on each of the following dimensions: 'benevolence', 'competence', and 'integrity'. As we did with the EFA, we identified all our observed variables as categorical data. The fit indices for the model are as follows: CFI = .988; TLI = .982; RMSEA = .102; AIC = 14,918; BIC = 15,153. This means that fit indices CFI and TLI indicate a very good model fit.

On the other hand, the RMSEA value of .102 is just above the .1 cut-off for 'mediocre fit'. Despite this violation, we believe that the overall fit of our model is good, because the value of universal cut-off points for the RMSEA index can be called into question. After careful empirical evaluation, Chen et al. (2008) found that universal cut-off points for RMSEA are questionable because 'the nature of this test is very different from the conventional hypothesis test such as the t-test, for which the relationship between the critical value α and the power of the test is known' (Chen et al., 2008: 490). As a result, they argue, it is important to use other goodness-of-fit measures to assert the global model fit. Furthermore, Breivik and Olsson (2001) found that RMSEA tends to favor complex models with many variables and constructs over models that are simpler and more parsimonious, which is the case in our relatively parsimonious (9 variables, 3 latent constructs) 'Citizen Trust in Government Organizations' (CTGO) scale. Therefore, multiple indices should be examined (Bentler, 2007; Chen et al., 2008), and a researcher must combine these indices with human judgment to ultimately decide about model fit. In conclusion, after examining the other indices which indicate excellent model fits – and the fact that the RMSEA value is only slightly above 'the standard' – we are confident that the current specifications of the model are acceptable nevertheless. The above conclusion is undergirded by the results we found examining single items in our model. All items loaded significantly on the latent variables ($p < .001$), and factor loadings were high (ranging from .809 to .926).

Next, we asserted the relationships between the three dimensions. All three dimensions are significantly correlated (see Figure 1). The correlation between perceived benevolence and competence is .772 ($p < .001$). The correlations between perceived integrity on the one hand and perceived benevolence and competence on the other are .887 and .765, respectively ($p < .001$). These correlations are fairly
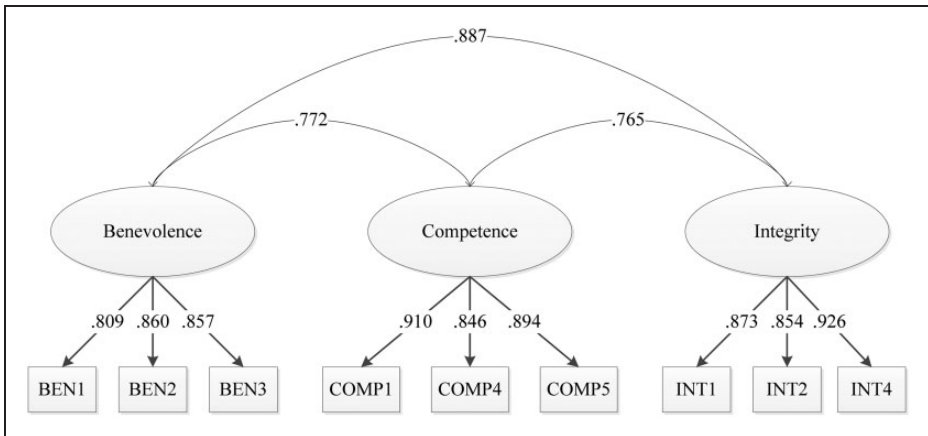
Figure 1. Factor structure trust scale.

high and may point toward uni-dimensionality. In order to assert this we also conducted a CFA in which we loaded all nine items measuring benevolence, competence and integrity on one factor. If the data indeed better fit a one-dimensional structure, then the fit indices should improve. However, fit indices drop well below the commonly accepted thresholds and thus indicate that the model fit is unacceptable (CFI = .947; TLI = .929; RMSEA = .205; AIC = 15,391; BIC = 15,611). This means that the multidimensional structure fits better with the data than the one-dimensional structure.

We also compared the fit of our three-dimensional model with three two-factor models. The fit indices of the two-dimensional model combining benevolence and competence in one dimension are as follows: CFI = .961; TLI = .946; RMSEA = .180; AIC = 15,260; BIC = 15,485. This indicates that our three-dimensional model fits the data better. The same holds when we compare our three-dimensional model with the two-factor model combining benevolence and integrity in one dimension (CFI = .982; TLI = .975; RMSEA = .121; AIC = 14,991; BIC = 15,216) and the two-factor model combining competence and integrity in one dimension (CFI = .955; TLI = .937; RMSEA = .194; AIC = 15,311; BIC = 15,536), which supports the assumption that trustworthiness is indeed a three-dimensional construct.

To further establish the reliability of our CTGO scale across different samples, we assessed the model fit on three sets of subsamples. First, we collected data from two different samples. The separate samples also show acceptable fit indices for the CFA (sample 1: CFI = .994; TLI = .991; RMSEA = .067; sample 2: CFI = .988; TLI = .981; RMSEA = .109). Factor loadings vary between .709 and .933 for sample 1 and between .838 and .914 for sample 2. Thus, the results of the CFAs support a three-dimensional scale comprised of 3 × 3 items across our two different samples. Second, because previous research suggests that educational level is an

important antecedent of trust (e.g. Cook and Gronke, 2005; Van Elsas, 2014), we conducted two CFAs to check whether our factor structure holds for both low-skilled and high-skilled citizens. Fit indices were as follows: CFI = .989; TLI = .984; RMSEA = .105 (low-skilled) and CFI = .987; TLI = .980; RMSEA = .108 (high-skilled). Further, factor loadings were very similar across groups. This indicates that our measure can be used to validly measure trust for both groups of citizens. Third, because previous research has indicated that political preference is significantly related to trust (Cook and Gronke, 2005; Orren, 1997), we conducted two CFAs to check whether the factor structure holds for both respondents who have a left-wing preference and those who have not. The fit indices were as follows: CFI = .985; TLI = .987; RMSEA = .111 (left-wing political preference) and CFI = .991; TLI = .987; RMSEA = .095 (others). Also, factor loadings were very similar across groups. As was the case with educational level, this indicates that our measure can be used to validly measure trust for both groups of citizens.

Finally, we assessed the reliability of the scales by examining the coefficients Cronbach's alphas. We were guided by Nunnally's (1978) suggestion that Cronbach's alpha should be at least .70 for acceptable reliability. The three sub-scales of trust all showed sufficient reliability. Cronbach's alphas were .831 for 'benevolence', .870 for 'competence' and .860 for 'integrity'. Again, these results are based on analyses performed on the complete dataset. Table 3 also displays Cronbach's alphas for the separate samples.

In summary, the results of our analyses show that the 3×3-item scale measuring trust is a reliable measure, and has a similar structure across different sets of subsamples (political leaning, educational degree).

Phase 3: Convergent validity

The goal of the third phase was to determine whether our multidimensional measure is significantly related to an often used measure of trust (convergent validity).

The one-item trust measure ('I have trust in XXX, when it concerns XXX') is significantly related to benevolence ($r = .779$, $p < .001$), competence ($r = .662$, $p < .001$), and integrity ($r = .678$, $p < .001$).

Second, we performed a regression analysis with the three dimensions of our trust measure as independent variables and a scale measuring trust in government in general as a dependent variable. The results show that all three dimensions are

Table 3. Cronbach's alphas

	Sample 1 ($n = 156$)	Sample 2 ($n = 835$)	Total dataset ($n = 991$)
Benevolence	.766	.843	.831
Competence	.852	.873	.870
Integrity	.820	.864	.860

significantly related to general trust in government. The standardized regression coefficients are as follows: $\beta = .246$, $p < .001$ (benevolence), $\beta = .137$. $p < .001$ (competence), and $\beta = .359$, $p < .001$ (integrity); 49.2 percent of the variance in general trust in government is explained by our three dimensions.

The results of phase 3 indicate that our multi-dimensional measure of trust is convergent with the one-item measure of trust and a scale measuring general trust in government. These findings therefore provide evidence of convergent validity for our trust scale. This means that our measure is correlated with a very relevant current one-item measure of trust in government and a scale measuring general trust in government. This provides evidence that our scale measures the theorized construct (i.e. trust) that it purports to measure.

Conclusion and discussion

The purpose of this study was to develop a valid and reliable scale to measure trust in a government organization at the meso level. Based on psychometric analyses, we excluded three of the initial items, resulting in a 3×3-item scale including perceived competence, benevolence, and integrity. The analyses show that this three-dimensional scale was psychometrically valid, and had a strong internal consistency. We demonstrated convergent validity between the three trust dimensions and an often used one-item measure of trust and a scale measuring general trust in government.

In addition, our scale measures trust in government organizations in a consistent way over various groups of respondents, because it is shown to have similar factor structures over the two subsamples with different types of respondents (students vs diverse sample). In addition, the measurement structure was also very similar when comparing a high and low educational group of respondents, which indicates that the scale measures the same dimensions for citizens regardless of their educational level. Interestingly, a recent scale analysis on political trust by Van Elsas (2014), using a different scaling technique, and a different sample and variables, comes to the same conclusion: the covariance structure of political trust is similar over educational level. Likewise, we found that the CTGO scale is valid for different groups of respondents, with regard to differences in educational level and political leaning.

This scale validation provided a first test, and we encourage other scholars to further validate the CTGO scale in other contexts. We think that the templates used in the scale allow for sufficient flexibility so that the usability of this scale extends beyond this particular domain (air pollution) and country (the Netherlands). However, to really establish the reliability and validity of the CGTO scale, future studies should test the scale in other cultural contexts, because the interpretation of the translated items may differ from culture to culture. Another way of further establishing the reliability of the scale using other templates is, for instance, by applying the items to other government organizations than municipal government. Templates could also be used to test the scale in other

policy contexts. Furthermore, we did not employ random sampling, but deliberately focused on two different subsamples to first test the stability of the scale. Further testing requires a test with a randomly selected sample.

Another issue that future research may look into is whether the CTGO scale can be extended with other dimensions of trust. In this study we only focused on the three most commonly distinguished dimensions of trustworthiness. However, broader models also include trusting intentions and behavior besides perceived trustworthiness. Although actual behavior is hard to measure by a questionnaire alone, intentions can be measured.

That said, the CTGO scale can help to further trust research in the field of public administration. Identifying specific objects of trust to determine the causes of trust in government organization more precisely is an important road ahead in public trust research (Van de Walle, 2004; Van de Walle et al., 2005). Assessing trust at a meso level, i.e. trust in a specific government organization, complements current models of trust in government at the macro level as measuring trust in government organizations allows us to use a multi-dimensional measure. Using multiple dimensions to measure trustworthiness of government organizations, such as with the CTGO scale, can reveal the antecedents and effects of trust more precisely.

Acknowledgment

Notes

1. Whether 'trust in government' at the macro level is multidimensional or not is subject to debate. However, because we focus on the meso level of trust we will not be involved in this debate. See Hooghe (2011) and Fisher et al. (2010) for a discussion on this issue.
2. It should be noted that 'distrust' is conceptually different from 'trust' or 'no trust'. Distrust is not the other end of high trust on the continuum. Trust and distrust are acknowledged to be distinct: a lack of trust does not necessarily imply distrust. Distrust is the presence of suspicion, whereas a lack of trust is merely the absence of trust (Cook and Gronke, 2005; Lewicki et al., 1998). Distrust is defined in terms of confident negative expectations regarding another's conduct, whereas trust refers to people's positive expectations. In this article we focus on the latter.

References

Ayres I and Braithwaite J (1992) *Responsive Regulation: Transcending the Deregulation Debate*. New York: Oxford University Press.

Bachmann R (2011) At the crossroads: Future directions in trust research. *Journal of Trust Research* 1(2): 203–213.

Bentler PM (1990) Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238–246.

Bentler PM (2007) On tests and indices for evaluating structural models. *Personality and Individual Differences* 42: 825–829.

Bouckaert G and van de Walle S (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* 69(3): 329–343.

Bovens Mark AP and Wille A (2008) Deciphering the Dutch drop: Ten explanations for decreasing political trust in the Netherlands. *International Review of Administrative Sciences* 74(2): 283–305.

Breivik E and Olsson UH (2001) Adding variables to improve fit: The effect of model size on fit assessment in LISREL. In: Cudeck R, Jöreskog KG, du Toit SHC and Sörbom D (eds) *Structural Equation Modeling: Present and Future : A Festschrift in Honor of Karl Jöreskog*. Lincolnwood, IL: Scientific Software International, pp. 169–194.

Browne MW and Cudeck R (1993) Alternative ways of assessing model fit. In: Bollen K and Long J (eds) *Testing Structural Equation Models*. Newbury Park: Sage, pp. 136–162.

Carnevale DG (1995) *Trustworthy Government: Leadership and Management Strategies for Building Trust and High Performance*. San Francisco: Jossey-Bass.

Chen F, Curran PJ, Bollen KA, Kirby J and Paxton P (2008) An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Social Methods Research* 36(4): 462–494.

Citrin J (1974) Comment: The political relevance of trust in government. *American Political Science Review* 68(3): 973–988.

Cook TE and Gronke P (2005) The skeptical American: Revisiting the meanings of trust in government and confidence in institutions. *Journal of Politics* 67(3): 784–803.

Cummings LL and Bromily P (1996) The organizational trust inventory: Development and validation. In: Kamer RM and Tyler TR (eds) *Trust in Organizations: Frontiers of Theory and Research*. Thousand Oaks: Sage, pp. 302–329.

Dalton R (2004) *Democratic Challenges, Democratic Choices: The Erosion of Political Support in Advanced Industrial Democracies*. Oxford University Press: Oxford.

DeVellis RF (2003) *Scale Development: Theory and Applications*. Thousand Oaks: Sage.

Fisher J, van Heerde J and Tucker A (2010) Does one trust judgement fit all? Linking theory and empirics. *The British Journal of Politics and International Relations* 12(2): 161–188.

Fukuyama F (1995) *Trust: The Social Virtues and the Creation of Prosperity*. New York: Simon & Schuster.

Granovetter M (1985) Economic action and social structure: The problem of embeddedness. *American Journal of Sociology* 91(3): 481–510.

Hardin R (1993) The street-level epistemology of trust. *Politics and Society* 21(4): 505–529.

Hardin R (2002) *Trust & Trustworthiness*. Sage Foundation Series on Trust. New York: Russell Sage Foundation, Russell.

Hetherington MJ (1998) The political relevance of political trust. *American Political Science Review* 92(4): 791–808.

Hibbing JR and Theiss-Morse E (2001) *Stealth Democracy: American's Beliefs about How Government Should Work*. Cambridge: Cambridge University Press.

Hooghe M (2011) Why there is basically only one form of political trust. *The British Journal of Politics and International Relations* 13(2): 269–275.

Kim S-E (2005) The role of trust in the modern administrative state: An integrative model. *Administration & Society* 37(5): 611–635.

Klingemann H-D, Volkens A, Budge I, Bara J and McDonald MD (2006) *Mapping Policy Preferences II: Estimates for Parties, Electors, and Governments in Eastern Europe, European Union and OECD 1990–2003*. Oxford: Oxford University Press.

Levi M and Stoker L (2000) Political trust and trustworthiness. *Annual Review of Political Science* 3: 375–407.

Lewicki RJ and Bunker BB (1995) Trust in relationships: A model of development and decline. In: Bunker JBB and Rubin JZ (eds) *Conflict, Cooperation, & Justice Essays Inspired by the Work of Morton Deutsch*. San Francisco: Jossey-Bass, pp. 133–173.

Lewicki RJ, McAllister DJ and Bies RJ (1998) *The Academy of Management Review* 23(3): 438–458.

Luhmann N (1979) *Trust and Power*. Chichester, UK: Wiley.

McAllister D (1995) Affect- and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal* 38(1): 24–59.

McEvily B and Tortoriello M (2011) Measuring trust in organisational research: Review and recommendations. *Journal of Trust Research* 1(1): 23–63.

McKnight DH, Choudhury V and Kacmar C (2002) Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research* 13(3): 334–359.

Mayer R and Davis JH (1999) The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology* 84(1): 123–136.

Mayer RC, Davis JH and Schoorman FD (1995) An integrative model of organizational trust. *Academy of Management Review* 20(3): 709–734.

Miller AH (1974) Political issues and trust in government: 1964–1970. *American Political Science Review* 68: 951–972.

Muthén L and Muthén B (2012) *Mplus User's Guide*. Los Angeles: Muthén & Muthén.

Norris P (1998) *Critical Citizens: Global Support for Democratic Governance*. Oxford: Oxford University Press.

Nunnally JC (1978) *Psychometric Theory*. New York: McGraw-Hill.

Nye JS, Zelikow PD and King DC (1997) *Why People Don't Trust Government*. Cambridge: Harvard University Press.

Nyhan RC (2000) Changing the paradigm: Trust and its role in public sector organizations. *The American Review of Public Administration* 30: 87–109.

Orren G (1997) Fall from grace: The public's loss of faith in government. In: Nye J, Zelikow P and King D (eds) *Why People Don't Trust Government*. Cambridge: Harvard University Press, pp. 77–108.

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Perry JL and Rainey HG (1988) The public–private distinction in organization theory: A critique and research strategy. *The Academy of Management Review* 13(2): 182–201.

Pina V, Torres L and Royo S (2007) Are ICTs improving transparency and accountability in the EU regional and local governments? An empirical study. *Public Administration* 85(2): 449–472.

Piotrowski SJ and Van Ryzin GG (2007) Citizen attitudes toward transparency in local government. *The American Review of Public Administration* 37(3): 306–323.

Poznyak D, Meuleman B, Abts K and Bishop GF (2014) Trust in American government: Longitudinal measurement equivalence in the ANES, 1964–2008. *Social Indicators Research* 118(2): 741–758.

Putnam R (2000) *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Rousseau D, Sitkin S, Burt R and Camerer C (1998) Not so different after all: A cross-discipline view of trust. *Academy of Management Review* 23(3): 393–404.

Tolbert C and Mossberger K (2006) The effects of e-government on trust and confidence in government. *Public Administration Review* 66(3): 354–369.

Van de Walle S (2004) *Perceptions of administrative performance: The key to trust in government?* Doctoral dissertation, University of Leuven, Belgium.

Van de Walle S, Kampen JK and Bouckaert G (2005) Deep impact for high-impact agencies? Assessing the role of bureaucratic encounters in evaluations of government. *Public Performance & Management Review* 28(4): 532–549.

Van de Walle S, van Roosbroek S and Bouckaert G (2008) Trust in the public sector: Is there any evidence for a long-term decline? *International Review of Administrative Sciences* 74(1): 47–64.

Van der Meer T (2010) In what we trust? A multi-level study into trust in parliament as an evaluation of state characteristics. *International Review of Administrative Sciences* 76(3): 517–536.

Van Elsas E (2014) Political trust as a rational attitude: A comparison of the nature of political trust across different levels of education. *Political Studies*, doi: 10.1111/1467-9248.12148.

Vigoda-Gadot E (2007) Citizens' perceptions of organizational politics and ethics in public administration: A five-year study of their relationship to satisfaction with services, trust in governance, and voice orientations. *Journal of Public Administration Research and Theory* 17: 285–305.

Vigoda-Gadot E, Beeri I, Birman-Shemesh T and Somech A (2007) Group-level organizational citizenship behavior in the education system: A scale reconstruction and validation. *Educational Administration Quarterly* 43(4): 462–493.

Vigoda-Gadot E, Shoham A and Vashdi RD (2010) Bridging bureaucracy and democracy in Europe: A comparative study of perceived managerial excellence, citizens' satisfaction, and trust in governance. *European Union Politics* 11(2): 289–308.

Yang K and Holzer M (2006) The performance–trust link: Implications for performance measurement. *Public Administration Review* 66(1): 114–126.

Zucker LG (1986) Production of trust: Institutional sources of economics structures, 1840–1920. In: Staw BM and Cummings LL (eds) *Research in Organizational Behavior*, Vol 8. Greenwich: JAI Press, pp. 53–122.

Stephan Grimmelikhuijsen is Assistant Professor at the Utrecht University School of Governance, the Netherlands. He is interested in issues of public sector transparency, citizen trust in government, e-government, red tape, and experimental public administration. Articles related to these topics have been published recently in various journals, such as *Journal of Public Administration Research and Theory*, *International Review of Administrative Science*, *Public Administration*, *Public Administration Review*, and *Government Information Quarterly*.

Eva Knies is Assistant Professor in Organizational Science and Human Resource Management (HRM) at the Utrecht University School of Governance, the Netherlands. Her research interests include public service performance, HRM in

the public sector, the role of middle and frontline managers, and scale development. Her research has been published in international journals, such as *Public Administration Review*, *Review of Public Personnel Administration*, *Social Policy & Administration*, *Human Resource Management Journal*, and *International Journal of Human Resource Management*.