Exploring Language Effects in Crosscultural Survey Research: Does the Language of Administration Affect Answers About Politics?

Zavala-Rojas, Diana

methods, data, analyses | Vol. 12(1), 2018, pp. 127-150 DOI: 10.12758/mda.2017.16

Exploring Language Effects in Cross-cultural Survey Research: Does the Language of Administration Affect Answers About Politics?

Diana Zavala-Rojas
Universitat Pompeu Fabra

Abstract

We study if the language of administration of a survey has an effect on the answers of bilingual respondents to questions measuring political dimensions. This is done in two steps. In the first we test whether the measurement instruments are equivalent for the same individual in two languages. After measurement invariance is established, we test if latent mean differences are significant across the two languages. We also test if the correlation of the same concept in two languages is equal to one or not. Results show evidence for language effects, the latent correlation is below one, although mean differences are not significant. We use data of the LISS migration panel in a within subject design, respondents answer a questionnaire twice first in Dutch and then in their (second) native language among Arabic, English, German, Papiamento and Turkish.

Keywords: language effects, bilingualism, measurement equivalence

Introduction

Target populations studied in large scale cross-national survey projects are linguistically diverse. In survey projects such as the European Social Survey, and the Survey of Health, Ageing and Retirement in Europe it is a common practice to translate questionnaires when at least 5% of the population is native speaker of a language (Dorer, 2012; SHARE, 2014), but little is known about the consequences and rationale behind this decision (Andreenkova, forthcoming 2018). In the present research we study if the language of administration of a survey influences the answers of *bilingual* respondents to questions measuring *political dimensions*. We define bilingual individuals in terms of language use, that is, individuals who have the ability to write, to speak, to read and to listen in two languages. Furthermore, they use both languages in their daily life: in their main activities such as work or school and with their friends and relatives (Grosjean, 2014).

Language effects in comparative survey research can have different forms; for instance, problematic translations can fail to reproduce the same stimuli across languages (Pennell et al., 2010; Davidov & De Beuckelaer, 2010), or the language of an interview usually activates cultural orientations driving individuals' responses (Luna, Ringberg & Peracchio, 2008; Peytcheva, forthcoming 2018). Language is a strong cultural carrier (Cohen, 2009) and bilingual individuals tend to live in mixed cultural environments. Cultural orientations may influence thoughts, cognitions and behaviour (Oyserman & Lee, 2008), and this in turn may affect the way respondents interpret and answer survey questions. Although translation issues have gained importance in comparative survey methodology, so far the effects of the language of administration on the responses to a questionnaire have received little attention in the field of survey research (exceptions are Peytcheva, forthcoming 2018; Elliot et al., 2012).

Research about language effects in the answers bilingual individuals give to measurement instruments has been conducted mainly in the fields of sociocultural psychology and psycholinguistics. In these two disciplines, even though diverse in methods and approaches, it has consistently been found that the language of administration of a questionnaire has an effect on the answers bilingual individuals give to cultural and self-identity items by activating specific cultural orientations linked to the language of the questionnaire (Chen & Bond, 2010, for a review; Chen, Benet-Martínez, & Ng, 2014). As the proportion of bilingual individuals is different across countries, the potential impact of this effect in cross-national survey research is unknown.

Direct correspondence to
Diana Zavala-Rojas, Universitat Pompeu Fabra (Spain)
E-mail: diana.zavala@upf.edu

To fill this gap, we have carried out a research project in which we test for language effects in a within-subject study of bilinguals in the Netherlands, a country with high linguistic diversity. Participants answered a questionnaire in Dutch and in their (other) native tongue: Arabic, English, German, Papiamento or Turkish. The first step is to test for measurement equivalence. Once equivalence is established, we test whether the correlation of a concept in two languages is equal to one. Third, we test if differences in latent means across languages were significant. The article proceeds as follows: In the next section, we introduce the mechanisms behind the effects of the language of administration on the answers to measurement instruments. Then, we introduce the operationalization of the concepts 'Trust and need of change in institutions' and 'Satisfaction and need of change in politics and the economy' and the models used to test for language effects. Afterwards, we explain the methodology we employ, that is, the procedures regarding the estimation and testing of the models. Next, we present the survey data we use. Finally we summarize the results and discuss the general findings.

Language Effects in Responses to Measurement Instruments

The mechanism behind the adaptation of responses as a function of the language in an interview can be explained by the theoretical frameworks of *acculturation* (Schwartz et al., 2014)with fully bilingual Hispanic participants from the Miami area, to investigate 2 sets of research questions. First, we sought to ascertain the extent to which measures of acculturation (Hispanic and U.S. practices, values, and identifications and *cultural frame switching* (*CFS*)*,* Honget al., 2000). As language is a strong cultural carrier (Cohen, 2009), individuals who master two languages may start an acculturation process, developing into a bicultural person (Grosjean, 2014) by internalizing to some extent the cultural attitudes and values attributable to the second language (Bond & Yang, 1982). Acculturation operates in three dimensions. The first is at the level of social behaviours or *practices*, such as cuisine preferences, language use and the choice of friends. The second is the acquisition of cultural *values*, for instance the importance of individualism versus collectivism. The third dimension is about identification: the attachment to a cultural, ethnic or national group (Schwartz et al., 2010).

CFS takes place when a person uses one system of cultural orientations instead of the other to react to specific social cognitions. This happens when cultural orientations are activated and become highly accessible in the mind of the person. Research has shown that the language of the interview can be a powerful activator of culture-specific mindsets in bilinguals, and individuals' answers to a questionnaire are adjusted accordingly (Bond & Yang, 1982; Chen, Benet-Martínez, & Ng,

2014; Chen & Bond, 2010; Luna et al., 2008; Schwartz et al., 2010, 2014; Yang & Bond 1980).

Previous research about language effects in bilingual individuals has been conducted in most cases with Asian subjects comparing their responses in Chinese and English languages, followed by research on the differences between Spanish and English in Hispanic communities in the United States. However, the dichotomies Chinese-Westerner or Hispanic-Westerner (where Western means English language or American culture) may be very specific cases. Both Chinese and Hispanic cultures emphasize collectivism as an archetypal trait, whereas preference for individualism is regarded as a Western archetype (Yoon, 2010). Respondents from highly communitarian cultures are more sensitized to contextual clues. They may assume that a certain type of culturally oriented response is expected (Lechuga, 2008). Moreover, the distance between Asian cultures and Western culture is perceived as very large (Minkov, 2007).

When language effects have been tested in other cultural contexts, findings have not been replicated completely. It remains unanswered to what extent language effects can be generalized to individuals of cultural backgrounds that are not Chinese or Hispanic. Other languages have been explored in fewer cases: for instance Arabic-French and Arabic-English (Botha, 1968), Afrikaans-English (Botha, 1970), Cebuano (Watkins & Gerong, 1999), French-English (Candell & Hulin, 1986), Greek-English (Richard & Toffoli, 2009; Triandis et al., 1965), Korean-English (Perunovic et al., 2007) and Russian-English (Marian & Neisser, 2000) and, to our knowledge, only one large scale study was conducted in more than 20 languages versus English (Harzing, 2006).

Language effects have been found consistently in responses to questionnaires about cultural dimensions (Benet-Martínez, Lee, & Leu, 2006; Bond & Yang, 1982; Harzing, 2005; Lechuga, 2008; Schwartz et al., 2014; Toffoli & Laroche, 2002; Triandis et al., 1965; Yang & Bond, 1980), personality perceptions (Chen et al., 2014; Chen & Bond, 2010; Ramírez-Esparza et al., 2006), feelings (Marian & Kaushanskaya, 2004; Perunovic et al., 2007), autobiographical memory (Marian & Neisser, 2000; Schrauf & Rubin, 2000), subjective evaluative ratings (Bond, 1985; Elliott et al., 2012; Pierson & Bond, 1982; Toffoli & Laroche, 2002) and self-relevant identity constructs (Dixon, 2007; Kemmelmeier & Cheng, 2004; Pierson & Bond,1982; Ross et al., 2002; Trafimow et al., 1997).

Luna et al., (2008) state that CFS only happens in bicultural bilinguals. The feelings and knowledge that monocultural bilinguals have associated to their second language does not affect how they see themselves. Consistently, several studies have found that language effects are mediated by individual characteristics related to *biculturalism*. Examples are the time in a lifespan and length of exposition to cultural practices of both cultures, and the extent they are perceived as compatible or oppositional; or the *language acquisition*: for instance in which setting the

languages were learned or the time of first exposition to each language (Benet-Martinez et al., 2002; Benet-Martínez & Haritatos, 2005; Dixon, 2007; Ji, Zhang, & Nisbett, 2004; Ross, Xun, & Wilson, 2002).

Benet-Martinez et. al., (2006) found out that biculturals' thinking about culture is more sophisticated than that of monocultural individuals. They are more experienced in dealing with cultural information because of their frequent CFS experiences. As a consequence, biculturals would have more complex cultural representations than monoculturals, but they were not expected to have complex representations in culturally neutral domains, such as geometric figures or landscapes. However with the exception of physical and mental health for which language effects did not emerge (Elliott et al., 2012; Peytcheva, 2008), *culturally neutral topics* have been tested in a few cases. Language effects have been studied in laboratory-settings on culturally neutral topics, being far too neutral, and of no relevance to social or political dimensions. Peytcheva (2018) argues that language effects would likely be present when the cultural specifics evoked by the language prompt cues of to what types of responses are socially accepted. Therefore, in the same survey interview, some items can be affected by language effects while for others this may not be the case.

There are several methodological limitations of most published research. The first is that language effects are tested by mean differences in composite scores of observed variables implicitly assuming that the measures are statistically equivalent across linguistic groups. Measurement equivalence is a prerequisite for cross-cultural comparison of models, relationships and means (Davidov et al., 2014; Meredith, 1993; Vandenberg & Lance, 2000). Before interpreting differences in responses, it is essential to test if the same measurement model on the relationship between indicators and latent variables holds in both languages. Only in few exceptions, measurement equivalence has been established prior to test for language effects in bilingual individuals (Candell & Hulin, 1986; Richard & Toffoli, 2009; Schwartz et al., 2014 test for measurement invariance and language effects).

A second methodological limitation in the analysis of language effects is that manifest variables are not measurement-error free. When differences in observed means have not been found to be significant, the conclusion has been that language effects are negligible. Only when full invariance is found, composite scores can be used directly. When partial invariance is found (Byrne, Shavelson, & Muthén, 1989), latent means should be used, composite scores are not adequate (Saris & Gallhofer, 2014, ch. 16).

A third limitation is that when mean scores are compared, it is, in general, not tested if the conceptual associations that individuals retrieve when they use one language or the other are the same, for instance when testing the strength of the correlation between a latent concept in one and the other language. Richard and Toffoli (2009) found that although the factorial structure of a construct (con-

figural invariance) and the way respondents answered (factor loadings invariance) were the same in Greek and English, the covariances between the latent variables were significantly different across languages. They argued that respondents had different conceptual associations in each language. A test where latent (or observed) mean differences are not significant does not rule out the possibility of language effects. It indicates that the distribution of the variable in the two languages is the same (equality in the location parameter) but that respondents can still have different conceptual associations in each language. In fact, evidence suggests that bilinguals may use different conceptual associations in each language, even in the cases where a literal translation exists (Ji et al., 2004; Luna et al., 2008). For instance, the language of an interview has been found to be a powerful activator of memories, individuals may retrieve auto-biographical experiences associated to the use of one language in consistency with the language of the interview. Marian and Neisser (2000) show that respondents interviewed in Russian (resp. English) remembered more experiences of their Russian-speaking (resp. English-speaking) period of their lives, depending on the language of the interview. For Hispanic bilinguals, autobiographical memories were encoded and retrieved in Spanish for events associated to the use of Spanish language, and in English for events in which English language was used (Schrauf & Rubin, 2000).

In our study we use a different approach to test for language effects. We use a specific application of a LISREL model (Jöreskog & Van Thillo, 1973), which we call in the following sections the *baseline model*. With this model, we test if the relationship across indicators and latent variables is the same in both languages. This is a test for measurement equivalence. Once it is established that the measurement model is equivalent, we are able to test structural relationships of latent variables in two languages. We test if two latent variables represent the same variable of interest by testing if its correlation is equal to one (Jöreskog, 1971; Saris, 1982a, 1982b). In other words, if two latent variables representing the same concepts in different languages had a very high correlation, the variables would be very similar across languages, nevertheless they would have a unique component indicating that they are not exactly the same.

Constructs, Survey Measures, and Models to be Tested

We test two concepts: "Political satisfaction" and "trust in institutions", both having a long tradition in political science and survey research. For these concepts we use a similar operationalization previously used in the European Social Survey Round 7 (European Social Survey 2015). Cultural orientations are known to affect political constructs (Inglehart, 1997; Crothers & Lockhart, 2000); thus, if the language of

the interview activates cultural orientations, bilingual individuals may score differently depending on the language of the interview.

In addition, we develop a measure of respondents‘ perception of political change. We operationalize the concept ‚political change‘ in a survey questionnaire following the three step procedure to formulate survey questions suggested by Saris and Gallhofer (2014). Appendix 1 shows the survey questions used in Model 1 and Model 2.

The first model we test is: ‚Trust and need of change in institutions‘ (Figure 1), consisting of two latent concepts. The first labelled ‚Trust‘ in Dutch institutions reflects the measures ‚trust in the parliament‘, ‚trust in the political parties‘, and ‚trust in the police‘. The second concept, ‚Need of change‘ reflects measures representing evaluative beliefs about the need of change in the way the aforementioned institutions work. Similarly, ‚Satisfaction and need of change in politics and the economy' (Figure 2) includes two concepts: ‚Satisfaction with politics‘ reflecting the indicators for ‚satisfaction with the economy‘, ‚satisfaction with the government‘, and ‚satisfaction with the democracy in the Netherlands‘. The concept ‚Need of change‘ reflects ‚the need of change in the economy‘, ‚the need of change in the way democracy works‘ and, ‚the need of change in the government‘. The left hand side of the figures represents the answers of the Dutch questionnaire, in the right hand side, the model corresponds to the same individuals answering in a second language (among Arabic, English, German, Papiamento and Turkish).

The η_j represent the jth latent variable; the y_{ij} is the ith observed variable for the jth latent trait and ε_{ij} are the disturbance terms; the λ_{ij} are the loadings; τ_{ij} are the intercepts and κ_j the latent means. It is assumed that the disturbance terms have a mean of zero and that they are uncorrelated with the latent variables. The disturbance terms are a combination of random errors and unique components. Thus, the unique components are correlated for the same observed variable in different languages denoted by $cov(\varepsilon_{11}, \varepsilon_{13})$, $cov(\varepsilon_{21}, \varepsilon_{23})$, ..., $cov(\varepsilon_{52}, \varepsilon_{54})$, $cov(\varepsilon_{62}, \varepsilon_{64})$. The other disturbance terms are assumed to be uncorrelated. The latent variables (η_j) are correlated with each other. In order to assign a scale to them, the loading of one observed variable is fixed to one, and the respective intercepts to zero (depicted with a dotted line in the pictures).

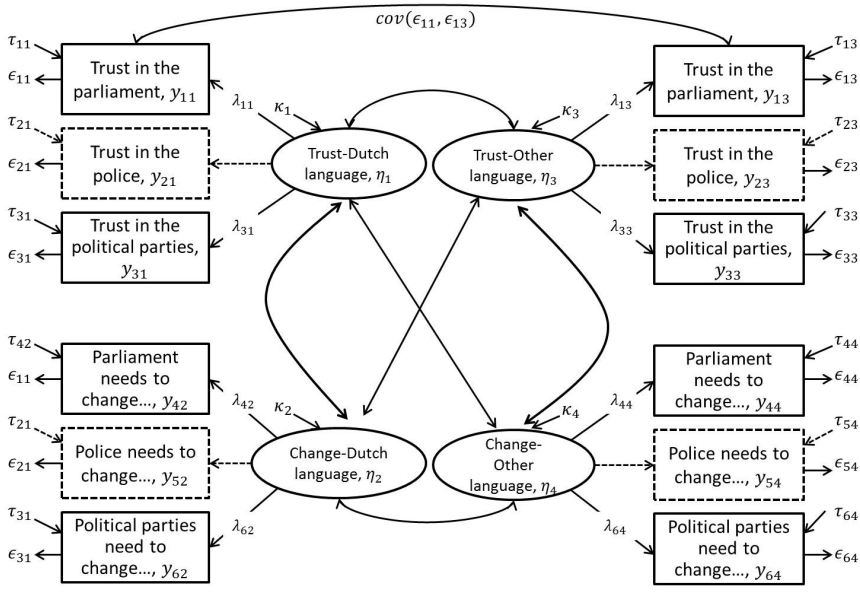

Figure 1 Model 1: Trust and need of change in institutions

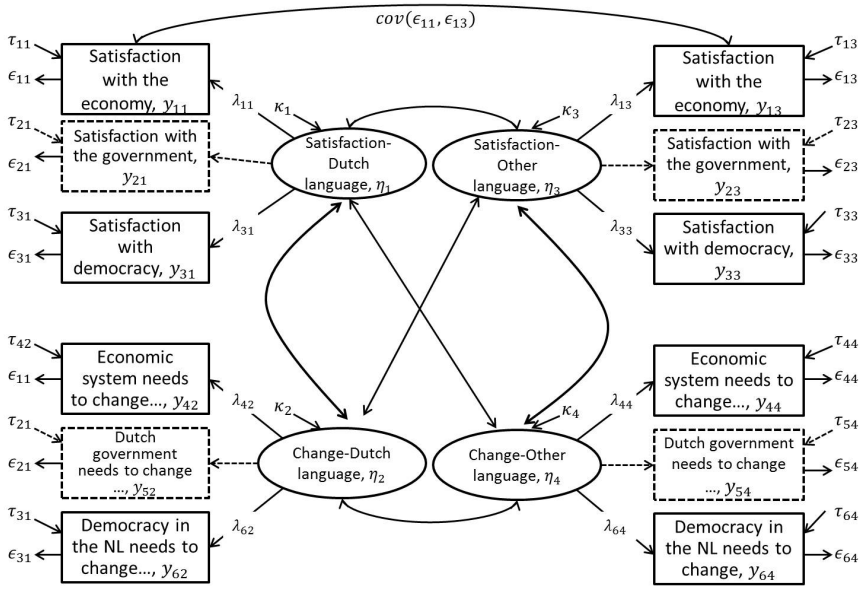

Figure 2 Model 2: ‚Satisfaction and need of change in politics and the economy

Method

We test for the measurement equivalence of measures answered in two languages by fitting a series of models starting with the baseline models shown in Figure 1 and Figure 2, and introducing consecutively equality constraints in the parameters (Davidov et al., 2014; Meredith, 1993; Vandenberg & Lance, 2000)[1]. First, we test that the same configuration of the factorial structure held in both languages. Second, the configural model is restricted to one where the factor loadings are constrained to be equal for the same manifest variable in a different language ($\lambda_{11} = \lambda_{13}$; $\lambda_{31} = \lambda_{33}$; $\lambda_{42} = \lambda_{44}$; $\lambda_{62} = \lambda_{64}$). When this restriction is not rejected, it is implied that comparisons of unstandardized relationships of observed variables across languages can be done. Thirdly, in addition to equivalence in the factor loadings, the intercepts are constrained to be equal ($\tau_{11} = \tau_{13}$; $\tau_{31} = \tau_{33}$; $\tau_{42} = \tau_{44}$; $\tau_{62} = \tau_{64}$). When the restriction in the intercepts is not rejected, it is implied that comparisons of means can also be done across languages.

Once equivalence in the measurement parameters is established, we further constrain the models to test first, whether the correlation between a construct in Dutch and in another language is equal to one ($\rho(\eta_1,\eta_3) = 1$; $\rho(\eta_2,\eta_4) = 1$). Failing this test is interpreted in the sense that the variables „reflect[ing] differences in conceptual associations among the true scores“ (Vandenberg, 2002, p. 142)and that it should be, Vandenberg and Lance elaborated on the importance of conducting tests of measurement invariance and proposed an integrative paradigm for conducting sequences of measurement invariance tests. Building on their platform, the current article addresses some of the shortcomings in our understanding of the analytical procedures. In particular, it points out the need to address (a and that they are not exactly the same, because they have a unique component in each language (Saris, 1982a). To estimate latent correlations and test whether or not they were one, two additional restrictions need to be imposed to the scalar models: the first is to fix the variances of the latent variables to one. The second, fixing the latent covariances of the same concepts in different languages to one. Using these constraints, the model estimates the matrix of standardized latent covariances, which are the latent correlations. Finally, we also test for invariance in the factor means ($\kappa_1 = \kappa_3$; $\kappa_2 = \kappa_4$). This restriction tests for differences between the two languages in the mean latent scores.

1 We estimated the models using Maximum likelihood estimation with ‘lavaan’ package for structural equation modeling (Rosseel 2012) in R3.1.2 statistical environment (R Core Team 2015). All reproducible scripts and the data for this article can be obtained from the author upon request.

Estimation and Testing of the Models

Goodness of fit (GoF) indices of structural equation modelling (SEM) are controversial (Cheung and Rensvold 2002). Commonly used fit criteria such as the Chi-square and the Root Mean Squared Error of Approximation (RMSEA) do not control for Type II error. We use the likelihood ratio test (LRT) in combination with the Judgement Rule (JRule) approach to test our models (Saris, Satorra, and Van der Veld 2009)[2]. The difference in the LRT indicates if the GoF is significantly worse for progressively more restrictive models. The JRule approach (Saris et al. 2009) identifies if fixed or constrained parameters are misspecified. A misspecification occurs if at each level of the equivalence tests specified in the previous section, a parameter has been given a fixed or constrained value, which is incorrect in the population of study (Hu and Bentler 1998). With this approach we can test directly for misspecifications in the models taking into account the power of the test for each misspecification. JRule works by combining knowledge of: (a) the size of the misspecification (expected parameter change); (b) the modification index, its impact on the fit if the parameter was freed in the model; and (c) the power of the test in detecting the misspecification[3]. Only when the modification index is significant and the power of the test low, the parameter is considered misspecified and freed in the models.

Saris et al. (2009) proposed a heuristic approach to choose the threshold for relevant differences. Following this recommendation, we detect standardized loading differences larger than 0.1, and intercept differences larger than 5% of the range of the response scales. As all measures had 11-point scales, this corresponds to intercept differences from 0.55. If a constrained parameter is misspecified according to JRule, it is freed and the null hypothesis of invariance in that restriction rejected. Once measurement equivalence is established, we set a threshold of 0.55 for differences in standardized latent means, which equals 5% of the items' scale. To test for equality of latent covariances/correlations, we restrict them to be equal between groups and test if this restriction was misspecified or not with a threshold of 0.10 for differences. For all decisions, we use a power of the test of 0.80.

Data

We conducted a two wave study between April and June 2013 in the Measurement and Experimentation in the Social Sciences (MESS) Immigrant Panel administered by CentERdata at Tilburg University, The Netherlands. The Immigrant Panel was

2 Appendix 2 reports global fit indexes.

3 The JRule approach for R is available in the 'miPowerFit' function, 'semTools' package (Pornprasertmanit et al. 2014).

a probability based online project in which researchers could submit proposals for fieldwork at no cost. Respondents were recruited based on stratified sampling using the population register as sampling frame. Participants were first and second generations of western and non-western origin of four major migration groups. They were provided with internet and a laptop to answer monthly surveys and received an economic incentive for each completed questionnaire.

The objective of Wave 1 was to select the languages to test for language effects in a within-subject design in Wave 2. The questionnaire included questions in Dutch about language use and knowledge, and questions about politics (see Appendix 1). All participants self-rated their ability in writing, listening, speaking and reading Dutch and their (second) native language in an 11-point scale (from 0 to 10). Wave 1 included 989 bilingual participants. They mentioned 74 languages as their native tongues. We selected the five languages in which respondents had the highest self-reported proficiency and the group was of at least 30 individuals: Arabic, English, German, Papiamento and Turkish. The source questionnaire was developed simultaneously in Dutch and English, translations into the other four languages were done by two independent translators, after which an adjudicator harmonized and decided upon the differences after discussing options with the translators. Questions were pretested with at least one person in each language. We based our procedure on the committee approach proposed by Harkness, Pennell, and Schoua-Glusberg (2004) for survey questionnaires, by involving a team in the translation process, although we simplify it due to budget restrictions.

In the second wave, the questionnaire was presented to 308 bilingual panel members, and it was fully completed by 255 respondents (83%). Due to the small number of individuals per language, the analysis was done within subjects, but it was not possible to separate the different linguistic groups. It was not possible to randomize the order of the languages, therefore, order effects may be present. The results presented in the next section are derived from this final sample size. Table 1 shows the mean and standard deviation of self-reported proficiency in both languages for participants who later on participated in Wave 2 (with the number of respondents by language and completion rates in parenthesis).

Although participants use their (other) native tongue in personal contexts such as at home and with their parents, at school and work, their predominant daily language is Dutch (Table 2). Turkish speakers have a balanced used of both languages with friends, and for German speakers, German language is less frequent in all aspects of life except with their parents.

Wave 2 consisted of three parts. In the first, individuals answered the core questions in Dutch. After that, they answered an unrelated questionnaire about different topics such as ideal body types, nature preservation, and King Willem-Alexander's succession. In the third part, they answered the core questions in Arabic, English, German, Papiamento or Turkish depending on the information they

provided in the first wave. Although memory effects cannot be excluded, they can be controlled for in the case of repetitions in survey interviews by asking other questions in between (Saris and van Meurs 1990).

Table 1 Mean self-reported proficiency in Dutch and target languages (standard deviation)

	Dutch				Target language			
Language group	Write	Read	Speak	Listen	Write	Read	Speak	Listen
English (n=104, 82.5%)	7.6 (2.4)	9.0 (1.4)	8.8 (1.5)	9.0 (1.5)	8.7 (1.7)	9.1 (1.4)	9.1 (1.2)	9.3 (1.3)
Papiamento (n=31, 86.1%)	7.1 (2.7)	8.5 (2.3)	8.6 (1.3)	8.9 (1.3)	6.3 (3.1)	7.4 (2.7)	8.5 (2.2)	8.8 (2.1)
Arabic (n=30, 83.3%)	5.9 (2.4)	7.0 (2.5)	7.0 (2.5)	7.4 (2.4)	7.8 (2.6)	8.2 (2.5)	8.8 (2.1)	9.0 (1.9)
German (n=35, 92.1%)	8.0 (1.8)	9.6 (0.8)	9.2 (1.3)	9.7 (0.7)	7.4 (2.4)	9.1 (1.3)	8.3 (2.1)	9.3 (1.1)
Turkish (n=55, 76.4%)	7.1 (2.5)	8.0 (2.2)	7.8 (2.1)	8.1 (2.0)	7.4 (2.5)	7.3 (2.6)	7.8 (2.2)	8.0 (2.0)

Table 2 Self-reported language use in Dutch and a second language

	Dutch language most frequently used... (%)				Second language most frequently used... (%)			
Language group of:	At work/ school	With friends	At home	With parents	At work/ school	With friends	At home	With parents
Arabic	92.6	56.7	40	0	3.7	33.3	53.3	88.2
English	70.2	81.7	51.9	43	29.8	16.3	47.1	70.7
German	85.7	97.1	85.7	26	8.6	2.9	11.4	47.3
Papiamento	100	70.9	54.7	14.2	--	25.5	45.4	71.2
Turkish	90.2	45.5	21.8	6	7.8	49.1	69.1	88

Note. Percentages adding Dutch and a second language for the same domain do not sum up to 100 when ‚other' language was reported as most used. For example, 56.7% of the Arabic speakers reported Dutch as their most frequently used language with friends, for 33.3% it was Arabic, and for 10% it was another language

Results

Equivalence in the Factorial Structure

Following the JRule test of local misspecifications, the baseline Model 1 (Trust and change in institutions) and Model 2 (Satisfaction and change in politics and the economy) are slightly modified. The p-value of the LRT is significant for the fit of the baseline model versus a model with some correlated errors (Table 3). In Model 1 we introduce two error covariances. The first is between the disturbance terms of the observed variable 'trust in the police' and 'need of change in the way the police works' ($cov(\varepsilon_{21}, \varepsilon_{52}) = cov(\varepsilon_{23}, \varepsilon_{54})$) and the second between 'trust in political parties' and 'need of change in the political parties' ($cov(\varepsilon_{31}, \varepsilon_{62}) = cov(\varepsilon_{33}, \varepsilon_{64})$). Both correlations are constrained to be equal across languages. In Model 2, we introduce three error covariances restricted to be equal between languages: 1) 'satisfaction with the economy' and 'need of change in the economy' $cov(\varepsilon_{11}, \varepsilon_{42}) = cov(\varepsilon_{13}, \varepsilon_{44})$, 'satisfaction with the government' $cov(\varepsilon_{21}, \varepsilon_{52}) = cov(\varepsilon_{23}, \varepsilon_{54})$ and 'need of change in the government' and 'satisfaction with the way democracy works in the NL' and 'change in the way democracy works in the NL' $cov(\varepsilon_{31}, \varepsilon_{33}) = cov(\varepsilon_{62}, \varepsilon_{64})$. Correlated errors improve the fit of the model and they are constrained to be equal across languages. Configural invariance is established because the same linear relationships exist between the indicators and the latent variables in both languages.

Equivalence in the Factor Loadings

Once we establish configural equivalence, we constrain the factor loadings to be equal across languages. As shown in Table 3, the LRT of the configural Model 1 and Model 2 are not significantly different from the restricted models. According to JRule this restriction is not misspecified. Therefore, equivalence in the factor loadings is established in both models.

Equivalence in the Intercepts

There are no significant misspecifications in the restricted intercepts. Furthermore, the LRT does not show that the fit was different between a model constraining loadings and a more restricting one which constrains intercepts. Full measurement equivalence is established in Model 1 and Model 2.

Table 3 Likelihood ratio test - Within subject measurement equivalence in Dutch and a second language

	Model 1					Model 2				
	DF	χ^2	$\Delta\chi^2$	ΔDF	$P(>)\chi^2$	DF	χ^2	$\Delta\chi^2$	ΔDF	$P(>)\chi^2$
Baseline model	42	209.9				42	232.8			
Baseline model + correlated errors	40	158.9	51.04	2	<0.001	39	172.4	60.42	3	<0.001
Invariance of loadings	44	165.0	6.06	4	0.19	43	175.4	2.996	4	0.558
Invariance of intercepts	48	170.8	5.82	4	0.21	47	179.8	4.415	4	0.353

Within-subject Structural Equivalence in Two Languages

Test for Cross-correlations Equal to One

We test whether the correlations between a latent variable in Dutch and the same latent variable in another language was equal to one, $\rho(\eta_1,\eta_3) = 1$; $\rho(\eta_2,\eta_4) = 1$). This is not the case in either Model 1 or in Model 2. Both the LRT and JRule indicate that this restriction should be rejected (Table 4). In Model 1, the correlation between 'trust' in Dutch and 'trust' in a second language is 0.78 ($\rho(\eta_1,\eta_3)$; and 0.64 between 'change' in Dutch and 'change' in a second language ($\rho(\eta_2,\eta_4)$. In Model 2, the correlation between the construct for 'satisfaction' in Dutch and 'satisfaction' in another language is not equal to one, but significantly lower (0.79) ($\rho(\eta_1,\eta_3)$. In the case of the CP 'change', the correlation between Dutch and a second language is of 0.71 ($\rho(\eta_2,\eta_4)$.

Test for Equal Factor Means

The fit Model 1 and Model 2 restricting latent means is not significantly different from the one restricting intercepts. According to JRule, we do not find misspecifications in the equality constraints of the latent means. In Model 2, the LRT shows that the fit of the model restricting latent means is significantly worse than the one which estimates the means without constraints. However, at the threshold level of 0.55 (5% of an 11-point scale), JRule does not indicate any significant differences in latent mean differences. When relaxing the threshold to detect deviations of 0.15 with a power of 0.80, JRule indicates that the equality constraints $\kappa_1 = \kappa_3$ and $\kappa_2 = \kappa_4$ are misspecified. The unstandardized estimate for the factor mean of 'satisfaction' is of 3.61 (se = 0.13) in Dutch language (κ_1) and 3.87 (se = 0.12) in the second language (κ_3). The unstandardized latent mean of 'change' is 6.98 (se = 0.12) in Dutch (κ_2) and 6.81 (se = 0.12) in the respondents' second language (κ_4). This

Table 4 Likelihood ratio test - Within subject differences in latent means and covariances

	Model 1					Model 2				
	DF	χ^2	$\Delta\chi^2$	ΔDF	$P(>)\chi^2$	DF	χ^2	$\Delta\chi^2$	ΔDF	$P(>)\chi^2$
Invariance of intercepts	48	170.8				47	179.8			
Correlations test	54	417.3	246.54	6	<0.001	54	495.4	315.55	7	<0.001
Latent means test	50	174.5	3.76	2	0.15	49	191	11.15	2	<0.004
Latent means with ‚satisfaction' mean free						48	182.6	2.75	1	0.09

result indicates that the mean scores of the underlying constructs that build Model 1 are significantly different in Dutch and in a second language for the same individual, however the difference is estimated around 1.5%. It is rather smaller than the threshold for mean differences established in Section 3.1.

Discussion and Conclusions

In the present study, we explore the effects of the language of the survey interview on the answers of bilingual respondents. Except for translation issues, the study of language effects on respondents' answers has received little attention in comparative survey methodology. As cross national comparative survey research expands to populations of study that are culturally diverse, measurement instruments are translated into more languages and more sampled individuals are themselves bilingual. This motivated the study of the potential effects that the language of the survey has on bilingual individuals. A limitation of this study is that the sample size is not large enough to divide the analysis by linguistic group in the bilingual sample, so further research is needed on specific cultural groups. A second limitation is that although the survey questions were repeated in the same survey interview, the true score for the same individuals using a different language may change with the passage of time or may include memory effects, thus changes may not only be due to switching to a different language. Nevertheless, this limitation is inherent to within-subject studies. A third limitation is that our findings cannot be generalized to other themes, they only hold for the tests in this study. Therefore, more research is needed to investigate the extent of language effects in different topics asked by the means of survey questionnaires.

Three specific research questions are addressed in the present study. The first is to investigate if language effects would emerge in bilingual individuals of cultural backgrounds different from those tested in the majority of published articles (Asian and Hispanic descendants). In our study, participants are bilinguals with Dutch as their main language. The second question is if language effects would emerge in political constructs: the reason for this research question is that so far, cultural and self-identity constructs have been explored in the literature rather than political topics. The third is to challenge the classical approach of testing for language effects comparing observed means of composite scores by testing whether the correlation of a latent variable in two languages is one.

In a first step we tested for within-subject measurement equivalence to confirm that our measures in two languages are invariant. Testing for measurement equivalence between languages has been seldom performed in past research, and it is a prerequisite for statistical comparison of survey items across cultures, languages and groups (Davidov et al., 2014). In a second step we tested for differences in latent correlations and means.

The first conclusion is that the measures we employ for the concepts in Model 1, 'Trust and need of change in institutions' and in Model 2, 'Satisfaction and need of change in politics and the economy' are statistically equivalent across languages. The second conclusion is that the language in a survey questionnaire affects to some extent the answers of bilingual respondents to political dimensions. We find, in both models, that the correlation between a latent variable measured by the same questions in Dutch and in a different language is not equal to one but significantly lower.

This is relevant to substantive research using these concepts because if the factors in Dutch and in another language have a very high correlation, the impact of each of them on a third variable will be difficult to distinguish. For instance, the larger the correlation between "political satisfaction" in Dutch and in another language, the more similar effect they have on "political participation". However, when the correlation is low, the association of "political satisfaction" with other variables of interest depends to some extent on the language of the survey measures. This would not be a problem if language effects were consistent across topics, but as we summarize in the literature review section, this is not the case.

Borrowing from cultural psychology the theoretical framework of *cultural frame switching* (CFS) (Hong et al., 2000), we interpret our results arguing that respondents made use of different conceptual associations in each language. As each language is associated with language specific cultural orientations, our results indicate that respondents shifted their cultural frame of reference when answering in the different languages.

However, factor mean differences did not emerge. This result indicates that language effects can be present even in the case when significant differences in

latent means do not emerge. Latent mean differences indicate a difference in the location of the parameters of the distribution of the latent variable[4].

Implications for Survey Methodology

Survey questions are measurement instruments of opinions. If the correlation between the same latent variable in two languages is not one, apparently it would follow that for certain topics, bilingual individuals are able to express two opinions, each triggered by cultural associations evoked by the language of the survey. The first implication of our findings for the design of surveys with multilingual samples is that the decision of the interview language should receive a more important role in the design of surveys. Andreenkova (2018) analyzes documentation on language choice in six large comparative survey projects finding out that information was very limited. The author concludes that more research needs to be done to design strategies for language allocation in bilingual populations, considering for instance, language usage and proficiency inquired from to the respondent at the beginning of the interview and using this information to select the language of the main interview. This would require interviewer training but also increasing survey agencies' awareness about the effects of the language of the interview.

Another possibility would be to give respondents two questionnaires in two different languages, as we did in this study, and average their opinion. From an operational point of view this solution is not optimal: For instance, it increases costs, increases cognitive burden on the respondent, increases the length of the interview and introduces potential memory effects. A third option (suggested in Richard & Toffoli, 2009) would be to randomize the questionnaires across languages. In a survey like the one presented in this study that would have meant that a random group of respondents would have answered in Dutch and another group in a second language. Although this option is statistically sound because differences across languages would cancel out, it is not operational in a comparative survey. The linguistic characteristics of the target population and of the individuals in the sampling frame are in general unknown before the data collection. Thus, the size of the random groups would be unknown as well. Moreover, functional bilingualism implies the combined abilities of writing, speaking, reading, and listening in two languages, and it also implies usage of both languages in their daily life (Grosjean, 2014). It does not imply that respondents feel fully comfortable answering certain topics in both languages.

Summing up, given the increasing evidence that language can affect responses to questionnaires in social and political surveys and in psychological instruments,

4 Very small significant latent means were found in Model 2, but they were well below the set threshold to consider them relevant.

providing an optimal solution on the choice of the language of the interview seems to be a clear aspect of comparative survey methodology that should receive more attention.

References

Andreenkova, A. (forthcoming 2018) How to Choose Interview Language in Different Countries. In: Johnson, T.P., Pennell, B.-E., Stoop, I. & Dorer, B. (Eds.). (forthcoming). Advances in Comparative Survey Methods: Multinational, Multiregional and Multicultural Contexts (3MC). Wiley Series in Survey Methodology. New York: John Wiley & Sons.

Benet-Martínez, V., & Haritatos, J. (2005). Bicultural Identity Integration (BII): Components and Psychosocial Antecedents. *Journal of Personality*, *73*(4), 1015–1050. https://doi.org/10.1111/j.1467-6494.2005.00337.x

Benet-Martínez, V., Lee, F., & Leu, J. (2006). Biculturalism and Cognitive Complexity: Expertise in Cultural Representations. *Journal of Cross-Cultural Psychology*, *37*(4), 386–407. https://doi.org/10.1177/0022022106288476

Benet-Martinez, V., Leu, J., Lee, F., & Morris, M. W. (2002). Negotiating Biculturalism: Cultural Frame Switching in Biculturals with Oppositional Versus Compatible Cultural Identities. *Journal of Cross-Cultural Psychology*, *33*(5), 492–516. https://doi.org/10.1177/0022022102033005005

Bond, M. H. (1985). Language as a Carrier of Ethnic Stereotypes in Hong Kong. *The Journal of Social Psychology*, *125*(1), 53–62. https://doi.org/10.1080/00224545.1985.9713508

Bond, M. H., & Yang, K.-S. (1982). Ethnic Affirmation Versus Cross-Cultural Accommodation: The Variable Impact of Questionnaire Language on Chinese Bilinguals from Hong Kong. *Journal of Cross-Cultural Psychology*, *13*(2), 169–185. https://doi.org/10.1177/0022002182013002003

Botha, E. (1968). Verbally Expressed Values of Bilinguals. *The Journal of Social Psychology*, *75*(2), 159–164. https://doi.org/10.1080/00224545.1968.9712488

Botha, E. (1970). The effect of language on values expressed by bilinguals. *Journal of Social Psychology*, *80*(2), 143. Retrieved from http://search.proquest.com/docview/1290717919?accountid=14708

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456.

Candell, G. L., & Hulin, C. L. (1986). Cross-Language and Cross-Cultural Comparisons in Scale Translations: Independent Sources of Information about Item Nonequivalence. *Journal of Cross-Cultural Psychology*, *17*(4), 417–440. https://doi.org/10.1177/0022002186017004003

Chen, S. X., Benet-Martínez, V., & Ng, J. C. K. (2014). Does Language Affect Personality Perception? A Functional Approach to Testing the Whorfian Hypothesis. *Journal of Personality*, *82*(2), 130–143. https://doi.org/10.1111/jopy.12040

Chen, S. X., & Bond, M. H. (2010). Two languages, two personalities? Examining language effects on the expression of personality in a bilingual context. *Personality and Social Psychology Bulletin*, *36*(11), 1514–1528.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen, A. B. (2009). Many forms of culture. *American Psychologist*, *64*(3), 194.

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75.

Dixon, D. J. (2007). The effects of language priming on independent and interdependent self-construal among Chinese university students currently studying English. *Current Research in Social Psychology*, *13*, 1–9.

Dorer, B. (2012). *Round 6 Translation Guidelines.* Mannheim.

Elliott, M. N., Edwards, W. S., Klein, D. J., & Heller, A. (2012). Differences by Survey Language and Mode among Chinese Respondents to a CAHPS Health Plan Survey. *Public Opinion Quarterly*, *76*(2), 238–264. https://doi.org/10.1093/poq/nfs020

European Social Survey. (2015). ESS Round 7: European Social Survey Round 7 Data. Bergen: Norwegian Social Science Data Services, Norway – Data Archive and distributor of ESS data.

Grosjean, F. (2014). Bicultural bilinguals. *International Journal of Bilingualism*. https://doi.org/10.1177/1367006914526297

Harkness, J. A., Pennell, B.-E., & Schoua-Glusberg, A. (2004). Survey Questionnaire Translation and Assessment. In *Methods for Testing and Evaluating Survey Questionnaires* (pp. 453–473). John Wiley & Sons, Inc. https://doi.org/10.1002/0471654728.ch22

Harzing, A.-W. (2005). Does the Use of English-language Questionnaires in Cross-national Research Obscure National Differences? *International Journal of Cross Cultural Management*, *5*(2), 213–224. https://doi.org/10.1177/1470595805054494

Harzing, A.-W. (2006). Response Styles in Cross-national Survey Research: A 26-country Study. *International Journal of Cross Cultural Management*, *6*(2), 243–266. https://doi.org/10.1177/1470595806066332

Hong, Y., Morris, M. W., Chiu, C., & Benet-Martínez, V. (2000). Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist*, *55*(7), 709–720. https://doi.org/10.1037/0003-066X.55.7.709

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453.

Ji, L., Zhang, Z., & Nisbett, R. E. (2004). Is It Culture or Is It Language? Examination of Language Effects in Cross-Cultural Research on Categorization. *Journal of Personality and Social Psychology*, *87*(1), 57–65. https://doi.org/10.1037/0022-3514.87.1.57

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426. https://doi.org/10.1007/BF02291366

Jöreskog, K. G., & Van Thillo, M. (1973). LISREL. Department of Statistics: University of Uppsala.

Kemmelmeier, M., & Cheng, B. Y.-M. (2004). Language and Self-Construal Priming: A Replication and Extension in a Hong Kong Sample. *Journal of Cross-Cultural Psychology*, *35*(6), 705–712. https://doi.org/10.1177/0022022104270112

Lechuga, J. (2008). Is Acculturation a Dynamic Construct?: The Influence of Method of Priming Culture on Acculturation. *Hispanic Journal of Behavioral Sciences*, *30*(3), 324–339. https://doi.org/10.1177/0739986308319570

Luna, D., Ringberg, T., & Peracchio, L. A. (2008). One Individual, Two Identities: Frame Switching among Biculturals. *Journal of Consumer Research*, *35*(2), 279–293.

Marian, V., & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language*, *51*(2), 190–201. https://doi.org/10.1016/j.jml.2004.04.003

Marian, V., & Neisser, U. (2000). Language-Dependent Recall of Autobiographical Memories. *Journal of Experimental Psychology: General*, *129*(3), 361–368. https://doi.org/10.1037/0096-3445.129.3.361

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Minkov, M. (2007). *What makes us different and similar: A new interpretation of the World Values Survey and other cross-cultural data*. Klasika i Stil Publishing House.

Oyserman, D., & Lee, S. W. S. (2008). Does culture influence what and how we think? Effects of priming individualism and collectivism. *Psychological Bulletin*, *134*(2), 311.

Pennell, B.-E., Harkness, J. A., Levenstein, R., & Quaglia, M. (2010). Challenges in Cross-National Data Collection. In *Survey Methods in Multinational, Multiregional, and Multicultural Contexts* (pp. 269–298). John Wiley & Sons, Inc. https://doi.org/10.1002/9780470609927.ch15

Perunovic, E., Wei, Q., Heller, D., & Rafaeli, E. (2007). Within-Person Changes in the Structure of Emotion: The Role of Cultural Identification and Language. *Psychological Science*, *18*(7), 607–613. https://doi.org/10.1111/j.1467-9280.2007.01947.x

Peytcheva, E. (2008). Language of administration as a cause of measurement error. In *AAPOR*. New Orleans.

Pierson, H. D., & Bond, M. H. (1982). How Do Chinese Bilinguals Respond To Variations of Interviewer Language and Ethnicity? *Journal of Language and Social Psychology*, *1*(2), 123–139. https://doi.org/10.1177/0261927X8200100203

R Core Team. (2015). R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from http://www.r-project.org/

Ramírez-Esparza, N., Gosling, S. D., Benet-Martínez, V., Potter, J. P., Pennebaker, J. W., & Ramírez-Esparza, J. W. (2006). Do bilinguals have two personalities? A special case of cultural frame switching. *Journal of Research in Personality*, *40*(2), 99–120.

Richard, M.-O., & Toffoli, R. (2009). Language influence in responses to questionnaires by bilingual respondents: A test of the Whorfian hypothesis. *Journal of Business Research*, *62*(10), 987–994. https://doi.org/10.1016/j.jbusres.2008.10.016

Ross, M., Xun, W. Q. E., & Wilson, A. E. (2002). Language and the Bicultural Self. *Personality and Social Psychology Bulletin*, *28*(8), 1040–1050. https://doi.org/10.1177/01461672022811003

Rosseel, Y. (2012). {lavaan}: An {R} Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36. Retrieved from http://www.jstatsoft.org/v48/i02/

Saris, W. E. (1982a). Different questions, different variables? In C. Fornell (Ed.), *A second generation of multivariate analysis. 2. Measurement and evaluation* (First, Vol. 2). New York: Praeger Publishers.

Saris, W. E. (1982b). Linear structural relations. In C. Fornell (Ed.), *A second generation of multivariate analysis: Methods* (First, Vol. 1). New York: Praeger Publishers.

Saris, W. E., Satorra, A., & Van der Veld, W. M. (2009). Testing Structural Equation Models or Detection of Misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. https://doi.org/10.1080/10705510903203433

Saris, W. E., & van Meurs, A. (1990). Evaluation of measurement instruments by meta-analysis of multitrait multi-method studies. *Proceedings, Amsterdam, 1989, Amsterdam: North Holland, 1990, Edited by Saris, W.E.; Meurs, A.van*, *1*. Retrieved from http://adsabs.harvard.edu/abs/1990emim.conf.....S

Schrauf, R. W., & Rubin, D. C. (2000). Internal languages of retrieval: The bilingual encoding of memories for the personal past. *Memory & Cognition*, *28*(4), 616–623.

Schwartz, S. J., Benet-Martínez, V., Knight, G. P., Unger, J. B., Zamboanga, B. L., Des Rosiers, S. E., … Szapocznik, J. (2014). Effects of language of assessment on the measurement of acculturation: Measurement equivalence and cultural frame switching. *Psychological Assessment*, *26*(1), 100–114. https://doi.org/http://psycnet.apa.org/doi/10.1037/a0034717

Schwartz, S. J., Unger, J. B., Zamboanga, B. L., & Szapocznik, J. (2010). Rethinking the concept of acculturation: Implications for theory and research. *2American Psychologist*, *65*(4), 237–251. https://doi.org/10.1037/a0019330

SHARE. (2014). The Survey of Health, Ageing and Retirement in Europe. Retrieved March 13, 2014, from http://www.share-project.org/

Toffoli, R., & Laroche, M. (2002). Cultural and language effects on Chinese bilinguals' and Canadians' responses to advertising. *International Journal of Advertising*, *21*(4), 505–524. https://doi.org/10.1080/02650487.2002.11104948

Trafimow, D., Silverman, E. S., Fan, R. M.-T., & Fun Law, J. S. (1997). The Effects of Language and Priming on the Relative Accessibility of the Private Self and the Collective Self. *Journal of Cross-Cultural Psychology*, *28*(1), 107–123. https://doi.org/10.1177/0022022197281007

Triandis, H. C., Davis, E. E., Vassiliou, V., & Nassiakou, M. (1965). *Some Methodological Problems Concerning Research Negotiations Between Monoinguals.*

Vandenberg, R. J. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures. *Organizational Research Methods*, *5*(2), 139–158. https://doi.org/10.1177/1094428102005002001

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Watkins, D., & Gerong, A. (1999). Language of Response and the Spontaneous Self-Concept: A Test of the Cultural Accommodation Hypothesis. *Journal of Cross-Cultural Psychology*, *30*(1), 115–121. https://doi.org/10.1177/0022022199030001007

Yang, K.-S., & Bond, M. H. (1980). Ethnic Affirmation by Chinese Bilinguals. *Journal of Cross-Cultural Psychology*, *11*(4), 411–425. https://doi.org/10.1177/0022022180114002

Yoon, K.-I. (2010). *Political culture of individualism and collectivism*. The University of Michigan.

Appendix 1

Survey questions administered in both languages

Model 1: Institutions: trust and change

Concept 1: Trust in institutions[5]

We will ask some questions about your level of trust in some institutions, 0 indicates complete distrust and 10 complete trust.

Overall, to what extent do you trust the parliament?
How much do you personally distrust or trust the police?
How much do you personally trust the political parties?

Complete distrust					Neither distrust nor trust					Complete trust
0	1	2	3	4	5	6	7	8	9	10

Concept: Need of change in the institutions

The next questions are about change in institutions, 0 indicates that the institution does not need to change the way it works and 10 indicates that it needs to completely change.

How much do you think that the Dutch parliament needs to change the way it works?

How much you think that the police needs to change the way it works to protect people like you?

To what extent do political parties need to change the way they work?

No need to change at all										Completely
0	1	2	3	4	5	6	7	8	9	10

5 The response scales were shown following each question, not in grids.

Model 2: Politics and the economy: satisfaction and change

Concept 1: Satisfaction with politics and the economy

Now we will ask you some questions about your satisfaction with some aspects of politics and the economy. Use a scale from 0 to 10 where 0 means you are completely dissatisfied and 10 means you are completely satisfied.

How satisfied are you with the present state of the economy in the Netherlands?
Overall, how satisfied are you with the way the Dutch government is doing its job?
And overall, how satisfied are you with the way democracy works in the Netherlands?

Completely dissatisfied					Neither dissatisfied nor satisfied					Completely satisfied
0	1	2	3	4	5	6	7	8	9	10

Concept 2: Need of change in politics and the economy

We will ask you about the level of change you think some aspects of in politics and the economy need, 0 indicates ‘there is no need at all to change’ and 10 is that ‘it needs to change completely’.

To what extent does the economic system in the Netherlands need to change?

Overall, to what extent does the Dutch government need to change the way it is doing its job?

To what extent does the way democracy work in the Netherlands needs to change?

Not need at all to change										Completely
0	1	2	3	4	5	6	7	8	9	10

Appendix 2

Global fit indexes of the models of models

Model 1. Trust and need of change in institutions

	DF	Chi-square	p-value	RMSEA	90 % confidence interval for RMSEA	CFI	SRMR
Baseline model	42	209.9	0	0.125	0.109, 0.142	0.917	0.071.
Baseline model + correlated errors	40	158.9	0	0. 108	0. 091, 0.126	0.941	0.060.
Factor loadings invariance	44	165	0	0.104	0. 087, 0.121	0.94	0.63.
Invariance of intercepts	48	170.8	0	0.1	0.084, 0.117	0.939	0.064.
Test of latent means differences	50	174.5	0	0.099	0. 083, 0.115	0.938	0.064.
Test of latent correlations = 1	54	417.3	0	0.162	0.148, 0.177	0.82	0.119.

Model 2. Satisfaction and need of change in politics and the economy

Baseline model	42	232.8	0	0.113	0.117, 0.150	0.916	0.072.
Baseline model + correlated errors	39	172.4	0	0. 116	0. 098, 0.134	0.941	0.070.
Factor loadings invariance	43	175.4	0	0.11	0. 093, 0.127	0.942	0.72.
Invariance of intercepts	47	179.8	0	0.105	0.089, 0.122	0.942	0.073.
Test of latent means differences	49	191	0	0.107	0. 091, 0.123	0.938	0.075.
Latent means test after freeing ‚sat' mean	48	182.6	0	0. 105	0.089, 0.121	0.941	0.073.
Test of latent correlations = 1	54	495.4	0	0.179	0.165, 0.194	0.806	0.239.