International Journal of Public Opinion Research Vol. 21 No. 3

doi:10.1093/ijpor/edp037

MEASUREMENT MODELS MATTER: IMPLICIT ASSUMPTIONS AND CROSS-NATIONAL RESEARCH

Tait R. Medina, Shawna N. Smith and J. Scott Long

ABSTRACT

The availability of cross-national survey data has grown exponentially in recent years. While much attention has been paid to increasing the comparability of indicators across countries, less has been done to increase the comparability of measurement models. This article examines the implicit assumptions of four different approaches to measurement modeling—summative scales, pooled exploratory factor analysis, multiple-group confirmatory factor analysis, and locally-conditioned factor analysis, and explores whether substantive conclusions in cross-national work can vary depending on the choice of measurement model. We find that results can vary by method and suggest that (i) the measurement modeling process itself be a critical part of cross-national research, and (ii) analysts be prepared to fully explain and defend measurement modeling decisions. A thorough understanding of the implicit assumptions of measurement modeling is required to avoid drawing conclusions that are little more than arbitrary.

During the past two decades there has been a stunning growth in cross-national survey research, due in no small part to the increasing availability of cross-national survey data. Surveys such as the World Values Survey (WVS), European Social Survey (ESS), International Social Survey Programme (ISSP), and the various Barometers (now in Europe, Latin America, Africa, and Asia) exemplify this trend, covering domains as diverse as citizenship, environment, family and gender roles, leisure time and sports, national identity, religion, social inequality and social networks across a wide range of countries (see Heath, Fisher & Smith, 2005 for a review).

The increasingly global reach of survey research has provided social scientists with new opportunities to pursue theory building and refinement through

This article was first submitted to *IJPOR* February 23, 2009. The final version was received July 7, 2009.

comparative analysis. Comparing attitudes and behaviors across countries and evaluating how associations between variables differ by context can clarify our understandings of causal mechanisms, relationships, and processes. But the very thing that makes cross-national survey research theoretically useful—comparisons across a broad array of contexts—also makes it vulnerable to methodological complexities and potential errors. Cross-national research requires careful consideration at both the conceptual and methodological levels. *Practitioners must be careful not to assume that the availability of cross-national survey data implies the appropriateness of cross-national comparisons.*

In spite of a long history of theoretical discussions about cross-national comparability, led by Verba (Almond & Verba, 1963; Verba, 1969), Smelser (1976), Prezerworski and Teune (1966, 1970) and others, significant issues remain regarding cross-national comparability, especially in terms of how concepts can be operationalized *equivalently* across countries. The purpose of our article is to highlight how implicit assumptions built into measurement models can affect conclusions from cross-national survey research in ways contrary to the intent of the researcher.

MEASUREMENT IN CROSS-NATIONAL RESEARCH

Measurement modeling consists of three distinct, yet interrelated components: *concepts* or abstract ideas drawn from theory; *indicators* or individual survey questions (also referred to here as "items"); and *constructs* or the combination of indicators into latent variables, which can take the form of factors, classes, or summative scales.

Concepts—Constructs—Indicators

Ideally, constructs are extracted from indicators in a way that reflects the concept of interest (Harkness, Mohler & Van de Vijver, 2003, p. 11). A main challenge in social science research is extracting constructs that are comparable across distinct groups. In cross-national research, such groups are countries that can be distinct in culture, social structure, and/or institutional arrangement.

A good deal of work in cross-national research has gone into ensuring that both *concepts* and *indicators* are comparable across countries. At the conceptual level, comparability often requires a level of abstraction. As Smelser (1976) explains, "to make concepts more widely comparative, then, is simultaneously to make them more abstract and inclusive" (p. 176). For example, Smelser suggests that the concept "administration" is superior to the concept "civil service," since the latter is linked to a bureaucratic administrative form not found in certain societies. At the indicator level, cross-national survey researchers have put a great deal of work into developing coordinated procedures that optimize reliability and validity across countries. Translation/back-translation procedures are commonly used, often including local experts, and some surveys have taken further action to ensure cross-national comparability. The ESS, for example, has relied upon multiple European experts to devise many of their indicators, and has also used "interactive" pre-testing to ensure indicator appropriateness. Similarly, the ISSP has a methodological research agenda investigating issues of translation, demographic comparability, and questionnaire design. Such research has advanced how survey researchers determine comparability, moving from privileging "sameness" to a better understanding of "functional equivalence". For example, a question that asks about religious advisors might need to be modified to include "traditional healers" in order to be culturally relevant (Pescosolido & Olafsdottir, 2008). In this way, non-literal translations may be necessary to create meaningful indicators in different contexts.

Comparability at the conceptual and indicator levels, however, *does not necessarily lead to comparability at the level of the construct*. Even if concepts and indicators have the same meaning across different contexts, the *link* between the two can vary. As such, even rigorous collection practices—such as those practiced by the ESS or the ISSP—do not eliminate the burden on the researcher to extract *equivalent conceptual meaning* from these questions. For example, in studying the stigma attached to individuals with mental illness, a question that asks whether a respondent would be willing to leave a child in the care of a person experiencing symptoms of mental illness could be interpreted similarly in all countries (*i.e.*, respondents from all countries understand, in the same way, what the question is asking), yet be a poor indicator of stigma in cultures where it is uncommon to leave a child with *any* caretaker other than immediate family, regardless of mental health status. In this case, an expression of willingness would not reflect lower stigma, but rather deviance from the cultural norms of child rearing. As such, the relationship between this indicator and others referencing stigma in domains such as work, friendship, and community would vary by cultural context, requiring a valid construct, then, that also varied by cultural context.

This article focuses on construct development in cross-national research first by discussing a variety of measurement models and highlighting the implicit assumptions and complications of each approach within cross-national work; and second by applying these approaches to the concept of national identity to evaluate whether measurement choice impacts substantive conclusions. That is, do conclusions differ depending on the measurement approach chosen?

The next section focuses on four approaches to cross-national measurement modeling. The first two approaches use a data set that combines all countries and assume *a priori* that constructs are invariant across countries: (i) pooled summative scales and (ii) pooled factor analysis. Two more context-sensitive approaches are then explored: (iii) multiple-group confirmatory factor analysis (MGCFA) which requires testing of cross-national invariance before meaningful comparison can proceed; and (iv) locally-conditioned factor analysis (LCFA) which allows for comparisons across all countries for which a valid local construct can be developed.

Section three applies these approaches to measuring national identity across multiple countries using the 1995 National Identity module of the ISSP. We have chosen this module because national identity is an example of a concept that both (i) is complex enough to require multiple indicators to measure and (ii) is affected by both local and global processes, thus illustrating the complexities of cross-national comparison. In addition, the ISSP ensures that the data have been subject to rigorous collection, mitigating measurement problems at the indicator-level.

The fourth section examines how choice of measurement model affects conclusions by focusing on the relationship between different measures of national identity and a set of predictor variables—for example, does the relationship between citizenship status and national identity vary by measurement model.

APPROACHES TO MEASUREMENT MODELING

A longstanding difficulty in the social sciences has been how to best measure social phenomena, such as democracy, trust, ideology, political efficacy, power, or well-being, given the absence of standardized units of measurement, such as mass, length, time or money (for a discussion of these issues, see Bulmer, 2001; Duncan, 1984). One approach to addressing this complexity has been to construct surveys with multiple indicators that are thought to proxy the concept of interest. Researchers use these indicators to develop constructs that reflect the concept. When this process of measurement modeling is carried out for cross-national comparison, assumptions regarding equivalency and comparability among countries are made, with specific assumptions varying by the type of measurement model. Accordingly, choosing among approaches is not a trivial methodological decision. Great care should be taken to determine which approach is best for the specific characteristics of the research. The remainder of this section provides an overview of four approaches to measurement modeling: (i) pooled summative scales; (ii) pooled factor analysis; (iii) MGCFA; and (iv) LCFA.

Pooled Measurement Models

Pooled approaches assume *a priori* that constructs are invariant across countries so that a single measurement model is sufficient for all countries. Two approaches to pooled measurement models are considered here. The "summative scale" is the most basic model, with indicators for a concept summed to create a score for each case. The summative scale assumes that each indicator measures a given concept equally and with the same reliability; that is, that indicators are "parallel" (Bollen, 1989, p. 208). A summative scale is equivalent to a factor model with all items having the same loadings and error variances. A summative scale can be created to represent one or multiple dimensions of a single concept, depending on theory and past research. For a single concept hypothesized to have one dimension, the summative scale would be expressed as follows:

$$y = cx_1 + cx_2 + \cdots + cx_p$$

where the summative scale score y weights each indicator x_k by c (usually $1/p$). To obtain the score for an individual i, we would need to know her values on $x_{i1}, \ldots, x_{ip}$.

Recent studies using pooled, summative scales in a cross-national framework include Fuwa (2004) and Batalova and Cohen (2002). In a cross-national framework the pooled summative scale approach assumes that both the form of the construct (number of dimensions) and the parallel weighting of indicators are invariant across countries. Each indicator has the same weight for all constructs in all countries, and the same indicators are used in all countries. There are many ways for this assumption to be invalid. For one, the concept could be unidimensional in some countries but not in others. In addition, a construct could be of similar form in all countries, but the link between indicators and construct could vary by country. Or, some indicators may not be meaningful within particular countries. Overall, the assumption that items are similarly linked to a construct not only in one country but across *all* countries is heroic and often unrealistic.

Factor analytic models relax some of the assumptions implicit in the summative scale and estimate the link between indicators and constructs. Exploratory factor analysis (EFA) is frequently used to determine the number and the nature of factors that account for the covariation between indicators. In this way, EFA is generally viewed as a theory-generating as opposed to theory-testing approach (Stevens, 1996). This technique allows the parameters to be freely estimated (aside from identification assumptions). As such, EFA allows multiple factors (*i.e.*, constructs) each with potentially different loading and error variances for each indicator. When applied to cross-national data, an EFA is run on all items assumed to contribute to the construct using the pooled data. Once the number of factors is decided, either based on theory or analysis of the data, researchers compute factor scores for each factor for each observation. These scores allow differential weights for each indicator. It should be noted, however, that some analysts choose to simply sum the items in each factor rather than compute factor scores—thus ignoring what is learned from the factor loadings (see, for example, Semyonov, Raijman & Gorodzeisky, 2006).

The pooled EFA model is more general than the summative scale approach since it allows for unequal contributions by each indicator to the construct. For a single factor score:

$$y = c_1x_1 + c_2x_2 + \cdots + c_px_p$$

where y is the factor score, the c_k's are factor score coefficients based on the estimated loadings, covariance of factors and covariance of indicators (Bollen, 1989, p. 305), and the x_k's are indicators as before.

Within cross-national work, although EFA relaxes some assumptions implicit in a summative scale, it continues to assume invariance of model form and coefficients across countries, including factor loadings, factor variances and covariances, and error variances. As with the summative scale approach, these assumptions can be invalid for many reasons. For example, an item might load highly on one factor in some countries but not others; or the model for the pooled data might extract two factors even though in some countries the concept requires three factors. *An EFA that fits the pooled data well does not necessarily indicate a good fit for each individual country.*

Pooled confirmatory factor analysis (CFA) allows the researcher to identify the model by specifying constraints on the model's parameters. Still, pooled CFA within a cross-national context is subject to the same limitations as EFA, and as such has the same potential problems.

MGCFA

While the pooled models discussed above *assume* that the measurement model is identical across populations, MGCFA allows analysts to *test* if this assumption is valid. Jöreskog (1971) presents a general confirmatory factor analytic approach that provides tests of the invariance of any parameter in the model across populations, including factor loadings, factor means, factor variances, factor covariances, and error variances. Muthén and Christoffersson (1981) have extended Jöreskog's approach to include categorical indicators.

In its most general form, a less restrictive model in which measurement parameters are free across groups is compared to a more restrictive model where measurement parameters are fixed to be equal across groups. If the fit of the more restrictive model is significantly worse than that of the less restrictive model, then invariance is rejected. It is broadly understood that meaningful cross-population comparisons that use scales generated from a measurement model require some degree of measurement invariance. Byrne, Shavelson, and Muthén (1989) argue that it is enough to have one equal factor loading per construct (plus the loading that is fixed to one for setting the scale) across populations to justify substantive comparisons. For further details about MGCFA, as well as discussions regarding the various degrees of invariance, see Muthén (1989); Meredith (1993); and Steenkamp and Baumgartner (1998).

In comparison to the pooled approaches, MGCFA does not start by assuming that a construct is equivalent in all contexts. From our perspective, this is a fundamental step in the right direction—after all, why assume measurement invariance *when invariance can be tested?* Therefore, we endorse MGCFA's proposition that the measurement form and parameters of each construct should be examined in each group (*i.e.*, country).

However, we do not believe that a statistical test of invariance solves the complicated problem of meaningful cross-national comparison. By setting measurement invariance as a prerequisite, MGCFA has the potential to exclude from comparison countries with measures that are similarly meaningful across contexts, yet not invariant. Just as cross-national researchers recognize that indicators may need non-literal translation to maximize the comparability between countries, constructs also may need non-literal translation. The next section outlines an alternative approach to measurement modeling that privileges *functional equivalence* rather than *invariance*, allowing for the non-literal translation of constructs across different contexts.

LCFA[1]

Smelser (1976) identified one of the central problems of cross-national research, the tension between the "universal" and the "particular", which he called the "double-tension":

> The search for appropriate comparative categories reflects a kind of double-tension: on the one hand, as just indicated, if the comparative analysis of dissimilar systems is desired, the investigator is under pressure to generate more abstract and inclusive concepts; on the other hand, the movement toward more abstract concepts creates a counterpressure for respecification of rules for identifying empirical indicators as they might manifest themselves in the dissimilar systems that have been encompassed by the more general categories. (p. 177)

To reconcile this tension at the indicator level, researchers have embraced the idea of functional equivalence. Functional equivalence emphasizes

[1] LCFA can be more broadly conceptualized as "group-specific" factor analysis, applying the same logic of functional equivalence to non-geographic distinctions such as gender, race/ethnicity or even time period.

"concordance of meaning" (Johnson, 1998) over literal translation. T.W. Smith (2003) notes that when developing comparable questions in cross-national surveys, "traditional rules mandate that methods should be identical across surveys, but the challenge is to identify cases in which methods identical on one level are not identical on other levels affecting measurement. In these instances, identical structure does not establish equivalency" (p. 82).

As Van de Vijver and Poortinga (1982) state, "concepts with functional equivalence are universal in a qualitative, although not necessarily a quantitative sense" (p. 390). In other words, the meanings are identical, but the model form and parameters are different. At the construct level, the universal and the particular can be integrated through the creation of constructs that meaningfully reflect the concept but are not necessarily identical across countries—what this section refers to as *locally-conditioned models*.

Functional equivalence is much harder to substantiate than measurement invariance since there is no statistical test for functional equivalence. Rather than relying on the Procrustean bed of statistical testing, functional equivalence is established through a critical process that requires both context-specific knowledge and engagement with the data.

Rigorous cross-national surveys have integrated functional equivalence at the indicator level into their practice, by using teams of experts to establish consistent meaning across countries. Ideally, such surveys would also integrate functional equivalence at the construct level, again employing experts in each country to develop measurement models for each concept under investigation. For a survey that includes fifteen countries, this might result in fifteen panels of experts arriving at fifteen different "locally-conditioned" measurement models that may or may not be *measurement invariant*, but would be *functionally equivalent*—and as such, directly comparable.

In the absence of functional equivalence being established by survey developers, however, the onus falls to the analyst to create such locally-conditioned models. To do this, the analyst utilizes country-specific expertise—for example, consultation with country-specific substantive literature or experts on a topic for a given country or region—to develop the most meaningful, appropriate and valid measurement model for each country. Just as factor analysis has both theory-testing (CFA) and theory-generating (EFA) variants, so too does LCFA. In other words, theory-testing LCFA would involve *a priori* hypothesizing based on pre-existing country-specific expertise. Alternatively, theory-generating LCFA would involve informed *post hoc* explanations of well-fitting, country-specific exploratory factor analyses.

Unlike approaches that rely on measurement invariance for establishing comparability, LCFA requires the analyst to establish the functional equivalence of models across countries, using the methods described above. As such, familiarity with each country is necessary for effective LCFA since such specific knowledge is required to bolster claims of functional equivalence. For this reasons, analysts should be selective in choosing countries to compare.

In practice, there are two requirements for determining country inclusion in any comparison. First, for each country, the model needs to fit the data well (see Bollen & Long, 1993, for a review of testing model fit). Second, the analyst needs to demonstrate, by using country-specific expertise, that each well-fitting model is a meaningful, appropriate and valid proxy for the concept of interest in the county at-hand. In other words, LCFA is not simply a mechanical application of factor analysis in different countries. Rather, LCFA requires that the analyst establish the functional equivalence of models by grounding such well-fitting models in context-specific knowledge. This is not an easy task, and complications will arise. Well-fitting models may not be found for some countries. Or, well-fitting models may not seem reasonable approximations for a concept given country-specific knowledge. In both cases, the source of the problem could be at the level of the construct, the indicator, and/or the concept. If the problem is with the construct, model re-specification, taking into account substantive knowledge of both the concept and the context, may be required. Further, problems could be at the indicator level. Does the question mean what you think it means? Investigating this type of problem may require returning to the original survey questionnaire in its country-specific language or reviewing the translation/back-translation process. Third, at the conceptual level, the concept may be too narrowly operationalized for the country of interest or make little sense for the country.

The locally-conditioned model, by privileging functional equivalence instead of factorial invariance, avoids the limiting assumptions of pooled models and MGCFA, but is not without limitation. First, model estimation for each country is more laborious than for a single pooled model since the locally-conditioned approach requires the integration of country-level theory and method. Second, significant differences in model form and dimensionality can make country comparisons more complicated. For example, how do you compare seven countries when two factors are sufficient in five countries, but three factors are required in the other two?

The four approaches to creating constructs that we have discussed carry different assumptions about the cross-national comparison of constructs. To this point, however, the discussion has been abstract, begging the question of whether different measurement models can lead to different conclusions. That is, are relationships between variables dependent upon measurement model choice? The next section applies the four approaches discussed above—summative scales, pooled EFA, MGCFA and LCFA—to the 1995 National Identity module of the ISSP to examine the impact of methodological choice.

AN EXAMPLE: NATIONAL IDENTITY

Although conceptual work on the social construction of national identity goes back to Anderson's (1983) seminal piece on "imagined communities", work by A.D. Smith (1991) has identified a more systematic basis of contextual difference in understandings of national identity. A.D. Smith makes the distinction between two types of nations and hence nationalisms: ethnic and civic. While the ethnic model views the nation in terms of genealogical descent, the civic model regards the nation as a community of citizens living under common laws and sharing a common civic culture. Jones and Smith (2001) extend A.D. Smith's (1991) conceptualization, arguing that national identity is better understood in terms of an ascribed dimension (related but not identical to ethnic identity) and a civic/voluntarist dimension.

Recent analyses of national identity often use the ISSP's module on National Identity to examine national identity in a comparative perspective. The National Identity module was first fielded in 1995 and partially replicated in 2003. The 1995 module was administered in 24 countries[2] to 30,894 respondents and includes seven items related to national identity (Zentralarchiv fuer Empirische Sozialforschung, 1998). See Table 1 for a description of these items, as well as sample Ns for all countries. Response rates were calculated for 11 of 23 countries, with exceptions due to quota sampling and/or sample substitution and/or a lack of supplied data, however not according to American Association for Public Opinion Research (AAPOR) standards (Park & Jowell, 1997). AAPOR standards were not implemented for the ISSP until the 2006 module (Gesis, *n.d.*).

In examining national identity across countries using the 1995 ISSP module, Jones and Smith (2001) argue that in most countries there are two dimensions of national identity: (i) an ascribed dimension comprised of items related to birth, residence, religion, and citizenship; and (ii) a civic/voluntarist dimension comprised of items related to national sentiment, a respect for institutions and laws, and fluency in the national language. The same dimensions are found in all countries except Germany, Spain, the Philippines, Bulgaria, and the Czech Republic. They write: "these factor patterns are repeated, with local variations, across most countries in the ISSP" (p. 54). While Jones and Smith paid little attention to these variations, such a statement epitomizes Smelser's "double tension" idea—while patterns persist across countries, so do local variations of potential importance. Accordingly, national identity provides an ideal site for examining the substantive impact of the four measurement model approaches discussed above.

[2] For the purposes of this article, we collapse East and West Germany into Germany, reducing the number of countries to 23.

TABLE 1 **Description and means of national identity items and covariates**

Variable	*Description*	*pooled*	*aus*[a]	*deu*	*gbr*	*usa*	*aut*	*hun*	*ita*	*irl*	*nld*	*nor*	*swe*	*cze*	*svn*	*pol*	*bgr*	*rus*	*nzl*	*can*	*phl*	*esp*	*lva*	*svk*	*jpn*
National Identity:																									
Some people say the following things are important for being truly (e.g. German). Others say they are not important. How important do you think each of the following is . . .[b]																									
Born	To have been born in (Country)	0.39	0.29	0.29	0.50	0.41	0.46	0.41	0.44	0.58	0.23	0.35	0.27	0.38	0.43	0.43	0.59	0.40	0.41	0.25	0.71	0.37	0.36	0.37	0.37
Citizen	To have (Country) citizenship	0.53	0.66	0.46	0.54	0.75	0.67	0.45	0.45	0.65	0.39	0.60	0.53	0.51	0.50	0.44	0.53	0.47	0.54	0.59	0.65	0.33	0.41	0.54	0.49
Residence	To have lived in (Country) for most of one's life	0.37	0.27	0.30	0.42	0.44	0.50	0.47	0.43	0.49	0.21	0.33	0.29	0.47	0.41	0.38	0.50	0.45	0.34	0.23	0.58	0.34	0.40	0.39	0.33
Language	To be able to speak (Country's language)	0.60	0.61	0.54	0.65	0.71	0.67	0.79	0.48	0.15	0.67	0.74	0.71	0.75	0.71	0.54	0.60	0.57	0.61	0.49	0.62	0.32	0.61	0.71	0.39
Religion	To be (Country's dominant religion[c])	0.20	0.15	0.16	0.22	0.39	0.32	0.20	0.26	0.32	0.03	0.10	0.08	0.11	0.17	0.27	0.45	0.17	0.15	0.15	0.57	0.18	0.14	0.12	0.11
Institutions	To respect (Country's) political institutions and laws	0.53	0.69	0.53	0.57	0.65	0.55	0.29	0.50	0.43	0.40	0.80	0.84	0.43	0.49	0.34	0.54	0.54	0.59	0.65	0.54	0.33	0.57	0.49	0.27
Sentiment	To feel (Country)	0.63	0.72	0.46	0.53	0.62	0.68	0.85	0.57	0.67	0.47	0.62	0.56	0.70	0.63	0.72	0.78	0.65	0.67	0.63	0.63	0.45	0.62	0.73	0.56
Covariates:																									
Age	Ranges from 14 to 98	45.08	48.84	47.10	46.82	44.50	46.32	47.78	42.91	45.98	43.73	43.13	45.00	42.95	42.79	47.35	49.08	44.74	46.45	41.52	39.62	44.81	47.15	41.16	45.75
Higher ed	1 = at least some university	0.24	0.36	0.16	0.26	0.30	0.04	0.11	0.16	0.25	0.24	0.26	0.26	0.17	0.18	0.13	0.20	0.40	0.30	0.38	0.17	0.13	0.48	0.10	0.28
Female	1 = female	0.53	0.50	0.47	0.60	0.56	0.55	0.57	0.52	0.51	0.54	0.50	0.51	0.49	0.56	0.55	0.52	0.55	0.53	0.51	0.50	0.52	0.61	0.52	0.54
Foreign	1 = spent most of childhood in another country	0.06	0.16	0.08	0.04	0.06	0.05	0.02	0.02	0.04	0.04	0.03	0.10	0.03	0.07	0.02	0.01	0.03	0.13	0.13	0.00	0.01	0.17	0.02	0.00
Parent citizen	1 = parents are citizens	0.91	0.73	0.94	0.94	0.90	0.93	0.98	0.98	0.96	0.95	0.95	0.86	0.97	0.90	0.98	0.98	0.99	0.79	0.77	0.99	0.99	0.64	0.96	1.00
Groups adapt	1 = it is better if different racial and ethnic groups adapt to larger society	0.45	0.75	0.35	0.66	0.43	0.47	0.35	0.56	0.54	0.59	0.64	0.68	0.41	0.36	0.30	0.39	0.15	0.56	0.53	0.43	0.42	0.24	0.34	0.11
N[d]		30870	2438	1894	1058	1367	1007	1000	1094	994	2089	1527	1296	1107	1036	1598	1105	1585	1041	1543	1200	1203	1044	1388	1256

[a]aus = Australia; deu = Germany; gbr = Great Britain; usa = United States; aut = Austria; hun = Hungary; ita = Italy; irl = Ireland; nld = Netherland; nor = Norway; swe = Sweden; cze: Czech Republic; svn = Slovenia; pol = Poland; bgr = Bulgaria; rus = Russia; nzl = New Zealand; can = Canada; phl = Philipinnes; esp = Spain; lva = Latvia; svk = Slovakia; jpn = Japan.

[b]The national identity items have a four-point response scale ranging from (1) very important to (4) not very important. For ease of presentation they are dichotomized here as 1 = very important and 0 = else. In subsequent analyses they are treated as ordinal and are coded as 1 = not very important; 2 = fairly important; 3 = important; and 4 = very important so that higher scores indicate higher levels of nationalism.

[c]In Norway this question specified Protestant Christian; in Italy, Ireland, Slovenia, Poland, the Philippines, and Spain it specified Catholic; in Russia, Orthodox; and in Japan, Buddhist or Shintoist. In all other countries, Christian was specified.

[d]24 cases from three countries were dropped due to duplicate id numbers.

Analysis Plan

We begin by using the four measurement approaches to create four sets of constructs for national identity. The first set includes summative scale scores corresponding to the two dimensions of national identity discussed above. The ascribed dimension adds the responses to the items Born, Residence, Religion and Citizenship and divides by the number of items with valid responses. The voluntarist dimension adds the responses for the items Institutions, Sentiment, and Language and again divides by the number of items with valid responses. The summative scales are standardized to zero mean and unit variance. The second set of constructs is factor scores from exploratory factor analyses for ordinal data (Muthén & Kaplan, 1985, 1992) using the pooled sample.

Third, we use MGCFA to evaluate measurement invariance. To do this, we first estimate a CFA based on Jones and Smith's specification of ascriptive and voluntarist identity (the same used to construct the summative scales) and choose those countries for which a well-fitting model can be found. We then test for both full and partial measurement invariance amongst those countries using the method outlined by Muthén and Asparouhov (2002) using the suggested delta parameterization. MGCFA begins by estimating an unrestricted model where the thresholds and factor loadings are free across countries, while the scale factors are fixed at 1 and factor means fixed at 0 in all countries. Next, a more restrictive model is estimated where the thresholds and factor loadings are constrained to be equal across countries, with the scale factors fixed at 1 in one country and free in the other countries, and factor means fixed at 0 in one country and free in the other countries. A difference of chi-square test is computed where a significant chi-square rejects the hypothesis that the constrained parameters are invariant across countries (*i.e.*, the equality constraints are violated). A non-significant chi-square test supports the invariance of the measures. If a lack of full measurement invariance is found, we test for partial measurement invariance. If full or partial measurement invariance is achieved across countries, factor scores are calculated providing our third set of constructs for national identity. For more information on MGCFA and related issues, see Muthén and Christoffersson (1981); Meredith (1993); and Millsap and Yun Tein (2004).

The fourth set of constructs is based on the LCFA approach. To maximize the number of countries included in our final comparisons across approaches, we begin by estimating models of national identity in each country using EFA. For those countries with reasonably well-fitting models, we then provide several illustrative examples of the literatures and expertise analysts could use in evaluating the meaningfulness of each model for each country.

We conclude by assessing whether measurement models matter, that is whether conclusions are dependent upon the measurement model chosen, using several predictor variables suggested in Jones and Smith's (2001) study of national identity. We do this in two ways. First, within any one country we examine whether the significance level of predictors varies across measurement approach in each country. Second, between any two countries, we use a fully interactive model to test if differences in effects vary by measurement approach. For example, does the effect of education on ascriptive national identity in the U.S.A. and Australia differ, and does that difference vary by measurement model?

Technical Details on Estimation

We use Mplus Version 5.2 for all factor analyses using robust weighted least squares for estimation (referred to as WLSMV in the Mplus documentation). Robust weighted least squares provides two advantages over weighted least squares: first, this method can deal with non-positive definite matrices by using a diagonal of the weight matrix (Muthén, du Toit, & Spisic, 1997). Second, simulation studies have shown that robust WLS is substantially more robust to modest violations of the continuous, normal latent process (Flora & Curran, 2004) assumed to determine each observed variable. By default, Mplus uses pairwise analysis with WLSMV. All analyses were rerun using listwise deletion with substantively similar results.

Constructing Measures

Fit statistics from the pooled EFA are presented in panel A of Table 2 and indicate that a two-factor solution is appropriate for the pooled data.

Assessing model fit has been much debated, and one suggestion has been for analysts to use multiple measures of fit (see Bollen & Long, 1993 for further discussion). We assess model fit using three criteria: the Tucker–Lewis Index (TLI) and the Comparative Fit Index (CFI), both comparative fit indices, and the Root-Mean-Square Error of Approximation (RMSEA), an error-of-approximation index. We use these fit criteria because their performance for categorical factor indicators has been assessed using simulation studies (Yu, 2002). For samples >250, Yu (2002) recommends the following cutoffs for acceptable model fit: TLI ≥ 0.95, CFI ≥ 0.96, and RMSEA ≤ 0.05. Browne and Cudeck (1993) provide further information on RMSEA values, categorizing RMSEA values >0.1 as "poor-fitting"; values in the range of 0.05–0.08 as "fair"; and values <0.05 as "close". For our purposes, a well-fitting model is required to meet Yu's recommendations for CFI and TLI, and also have a non-"poor-fitting" RMSEA (<0.1). Chi-square tests are also presented, but not used to determine fit since this statistic is sample size dependent and often too sensitive to trivially small misspecifications (Bollen & Long, 1993, p. 6).

TABLE 2 **Fit statistics from EFA for one- and two-factor models**

	Factors	χ^2	*df*	*P-value*	*CFI*	*TLI*	*RMSEA*
Panel A: Pooled EFA							
Pooled (30,586)	2	826.71	8	<.001	0.99	0.99	0.06
Panel B: Locally Conditioned EFA							
Australia (2,419)	2	22.59	7	.002	1.00	1.00	0.03
Great Britain (1,039)	2	19.81	7	.006	1.00	1.00	0.04
Netherlands (2,085)	2	35.14	8	<.001	0.99	0.99	0.04
Sweden (1,294)	2	23.95	8	.002	0.99	0.99	0.04
United States (1,350)	2	24.96	7	<.001	1.00	1.00	0.04
Austria (1,006)	2	31.32	8	<.001	0.99	0.99	0.05
Hungary (994)	2	25.36	7	<.001	0.99	0.99	0.05
New Zealand (1,020)	2	22.83	7	.002	0.99	0.99	0.05
Latvia (1,025)	2	32.63	7	<.001	0.99	0.98	0.06
Slovakia (1,386)	2	41.48	7	<.001	0.99	0.97	0.06
Slovenia (1,030)	2	33.68	7	<.001	0.99	0.99	0.06
Ireland (993)	2	43.92	7	<.001	0.99	0.98	0.07
Norway (1,516)	2	61.85	8	<.001	0.98	0.97	0.07
Canada (1,529)	2	74.90	7	<.001	0.98	0.97	0.08
Philippines (1,200)	2	71.35	7	<.001	0.99	0.99	0.09
Poland (1,568)	2	91.91	7	<.001	0.99	0.99	0.09
Bulgaria (1,068)	*2*	*80.10*	*7*	*<.001*	*0.98*	*0.98*	*0.10*
Italy (1,093)	*2*	*102.24*	*7*	*<.001*	*0.97*	*0.96*	*0.11*
Panel C							
Germany (1873)	*1*	*227.60*	*12*	*<.001*	*0.96*	*0.98*	*0.10*
Japan (1,245)	*1*	*172.93*	*12*	*<.001*	*0.96*	*0.97*	*0.10*
Russia (1,555)	*1*	*240.61*	*12*	*<.001*	*0.93*	*0.93*	*0.11*
Czech Republic (1,102)	*1*	*181.03*	*12*	*<.001*	*0.93*	*0.95*	*0.11*
Spain (1,196)	*1*	*398.93*	*11*	*<.001*	*0.95*	*0.98*	*0.17*

Note: Non-missing *N* in parentheses. Panel B includes all countries for which a two-factor model could be estimated. Panel C includes countries for which only a one-factor model could be estimated due to either non-convergence or negative residual variances. Italicized countries are excluded from further analyses due to poor-fit and/or non-convergence, and/or negative residual variances.

The geomin-rotated factor loadings for the pooled EFA are given in panel A of Table 3. Because a matrix of correlations can be factored in an infinite number of ways, rotation is essential for determining a factor form that is substantively meaningful. While rotation does not affect the fit of the data, it does change the interpretation. Geomin rotation, an oblique rotation, is used because we think that correlated factors are theoretically more plausible and this type of rotation is particularly good at minimizing the number of substantial loadings that appear on more than one factor (Browne, 2001). In practice, analysts will have to decide if this rotation method fits their theoretical needs. In examining patterns of factor loadings, we consider a loading on a factor to be "strong" if it is over 0.8 and twice its loading on the other

TABLE 3 **Geomin-rotated factor loadings from EFA for countries with well-fitting, two-factor solution**

	Factor	*Born*	*Residence*	*Citizen*	*Religion*	*Language*	*Institutions*	*Sentiment*	*Corr.*
Panel A: Pooled EFA									
Pooled	1	**0.93**	0.51	0.34	0.42	0.02	−0.33	0.01	
	2	0.00	0.36	0.50	0.21	*0.66*	*0.79*	*0.65*	0.60
Panel B: Locally conditioned EFA									
Australia	1	**0.90**	**0.82**	0.21	0.37	0.28	−0.36	0.00	
	2	−0.01	0.09	*0.57*	0.25	0.49	*0.69*	*0.61*	0.43
Great Britain	1	**0.99**	0.47	0.26	0.21	−0.01	−0.13	0.01	
	2	0.00	0.44	*0.63*	*0.50*	*0.78*	*0.72*	*0.75*	0.60
Netherlands	1	**0.92**	0.41	0.53	0.34	0.00	−0.33	0.08	
	2	0.00	0.42	0.35	0.12	**0.77**	*0.58*	0.49	0.48
Sweden	1	**0.89**	*0.65*	0.40	*0.52*	0.02	−0.31	0.32	
	2	−0.01	0.23	0.42	0.05	**0.80**	*0.74*	0.32	0.502
United States	1	**0.92**	**0.90**	0.56	*0.63*	0.51	−0.01	0.30	
	2	−0.07	0.00	0.37	0.12	0.37	*0.76*	0.54	0.38
Austria	1	**0.94**	*0.79*	**0.83**	*0.64*	*0.59*	0.00	0.35	
	2	−0.18	0.09	0.00	0.03	0.21	*0.63*	0.61	0.53
Hungary	1	*0.78*	*0.67*	**0.88**	*0.54*	0.00	0.41	−0.05	
	2	−0.01	0.25	0.00	−0.09	**0.94**	0.05	*0.50*	0.48
New Zealand	1	**0.93**	*0.65*	0.40	0.16	0.12	−0.32	−0.01	
	2	0.00	0.25	0.50	0.45	*0.58*	*0.67*	*0.52*	0.33
Latvia	1	**0.83**	0.51	**0.82**	0.33	0.46	−0.04	0.01	
	2	−0.04	0.29	0.00	0.10	0.41	*0.79*	**0.86**	0.34
Slovakia	1	**0.85**	0.49	0.61	0.32	0.29	0.01	−0.04	
	2	−0.01	0.38	0.31	0.16	0.55	*0.57*	*0.74*	0.32
Slovenia	1	*0.79*	0.50	*0.58*	*0.72*	−0.01	0.09	0.06	
	2	0.00	0.37	0.25	−0.15	**0.85**	*0.52*	*0.69*	0.61
Ireland	1	**0.92**	0.40	*0.62*	0.19	0.08	−0.13	0.00	
	2	0.00	0.38	0.28	*0.53*	*0.55*	*0.62*	*0.72*	0.66
Norway	1	**0.95**	0.45	0.34	0.43	0.00	−0.27	0.12	
	2	0.00	0.40	0.50	0.23	*0.74*	*0.66*	*0.51*	0.53
Canada	1	**0.96**	**0.83**	0.53	*0.64*	0.33	−0.01	0.25	
	2	−0.19	0.01	0.42	0.00	0.14	**0.81**	*0.57*	0.22
Philippines	1	*0.79*	**1.03**	*0.66*	0.43	*0.72*	0.00	0.38	
	2	0.01	−0.21	0.21	0.20	0.18	**0.86**	*0.55*	0.81
Poland	1	**1.01**	*0.66*	**0.88**	*0.60*	0.45	0.17	0.00	
	2	−0.24	0.27	0.01	0.01	0.50	0.46	**0.82**	0.56

Note: Bold numbers indicate "strong" factor loadings above 0.80 and twice the other factor loading. Italicized numbers indicate "moderate" factor loadings between 0.50 and 0.80 and twice the other factor loading.

factor, and "moderate" if it is over 0.5 and twice its loading on the other factor. This ad-hoc criteria points to an item that loads essentially on one factor.

The pattern of factor loadings for the pooled data fits generally with Jones and Smith's (2001) ascriptive-voluntarist conceptualization of national identity. Born loads strongly on factor 1, while Language, Institutions and Sentiment all load moderately or strongly on factor 2.

As discussed prior, the pooled EFA assumes that the same factor structure fits equally well in all countries. Next we use MGCFA to test this assumption.

TABLE 4 **Test of measurement invariance among countries with configural invariance (Germany, Austria, Hungary, Ireland, Norway, and Slovenia)**

	χ^2 *difference*	*df*	*P-value*
Full Invariance			
Unconstrained versus Constrained	2004.795	61	<.001
Partial Invariance			
Unconstrained versus Constrain only			
Born	23.143	8	.003
Residence	44.683	8	<.001
Citizen	19.675	8	.012
Religion	49.483	7	<.001
Language	25.347	8	.001
Institutions	30.274	7	<.001
Sentiment	69.925	8	<.001

Note: A *P*-value $\leq$.05 indicates a lack of measurement invariance.

First we check for "configural" invariance—the least-restrictive level of measurement invariance (Thurstone, 1947; Steenkamp & Baumgartner, 1998)—by estimating a CFA with the same pattern of zero and non-zero loadings in each of our 24 countries. As with the summative scale, we draw on Jones and Smith's (2001) research and estimate a two-factor model where factor 1 loads on Born, Citizen, Religion, and Residence and factor 2 loads on Language, Institutions, and Sentiment. CFA results indicate that configural invariance is supported for six of the 24 countries in the sample—that is, the same model fit the data well in these six countries. Next we estimate a baseline model where factor loadings and thresholds are free to vary across countries. For all models, we set the metric of the factor by setting the factor variance to 1 (for more on identification assumptions in MGCFA see Cheung & Rensvold, 1999; Millsap & Yun Tein, 2004). The unconstrained model provides a baseline chi-square statistic to compare with the constrained model. The constrained model, which holds the factor loadings and thresholds constant across all six countries, is estimated next. A chi-square difference test is used to compare the unconstrained and constrained model. For models estimated using the WLSMV, the conventional approach of taking the difference between the chi-square values and degrees of freedom is not appropriate because the chi-square difference is not distributed as chi-square. For this reason, difference testing in Mplus is done using the "diftest" option (Mplus Users' Guide). No other fit statistics are used to evaluate measurement invariance, as simulation studies have shown that measures such as the CFI may be problematic when used with WLSMV (French & Finch, 2006). Table 4 presents this omnibus chi-square difference test of full invariance. The chi-square difference test indicates that the constrained model fits the data significantly worse than the unconstrained model, indicating that full measurement invariance does not hold.

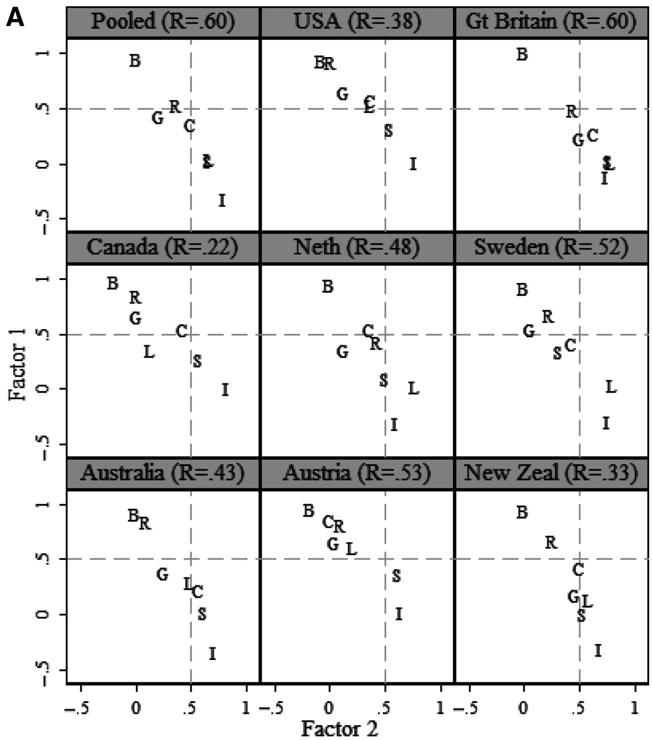
Next we examine partial measurement invariance. Following Byrne et al. (1989), we define partial measurement invariance as having at least one item per factor that is invariant. For identification, we fix the factor variance to 1 in all models. This allows us to estimate the factor loadings for each item separately, and then test each partially constrained model against the unconstrained model. Table 4 presents the chi-square difference tests between the baseline model and a series of models that constrain one item at a time to invariance across all six countries. A significant chi-square test indicates that the indicator is not invariant across countries. As the table shows, none of the individual items are invariant.[3] Since MGCFA argues that meaningful cross-group comparisons cannot be made without at least partial measurement invariance, this third measure of national identity is not constructed.

Fit statistics for the locally conditioned EFAs are presented in panels B and C of Table 2. Countries in each panel are ordered according to model fit, from best to worst. The first 16 countries of panel B (Australia through Poland) fit our criteria for a well-fitting model; the last two (Bulgaria and Italy) do not. Panel C presents the countries for which a reasonable two-factor solution could be not estimated. For Russia, convergence of a two-factor solution was not obtained even with 1,000 iterations. For the remaining four countries, a two-factor solution produced a negative error variance on the item Born. For purposes of illustration, we are dropping the seven countries italicized in Table 2 from further analyses. In a serious substantive analysis, however, one would need to consider why the data do not fit these models well and whether there may be problems at the concept and/or indicator levels.

The geomin-rotated factor loadings for the well-fitting models are presented in panel B of Table 3. To begin, we evaluate whether the pattern of factor loadings, *prima facie*, reflect the ascriptive-voluntarist conceptualization of national identity, where being born in a country (Born) is a strong indicator of ascribed identity and respecting a country's institutions and laws (Institutions) and national sentiment (Sentiment) are strong indicators of voluntarist identity. The remaining items—Citizenship, Residence, Language and Religion—could, depending on national context, be more or less related to each dimension. For example, Jones and Smith (2001) suggest that language is a "facilitator of civic virtue" (p. 106) and, as such, the saliency of language-as-facilitator is likely to depend on how isolated the country's national

[3] Simulation studies have found that when using the WLSMV estimator the power for detecting difference for a single variable is low and conclusions of invariance may therefore be erroneous. In other words, a finding of measurement non-invariance for all items tested individually provides further strong evidence of overall measurement non-invariance (French & Finch, 2006).

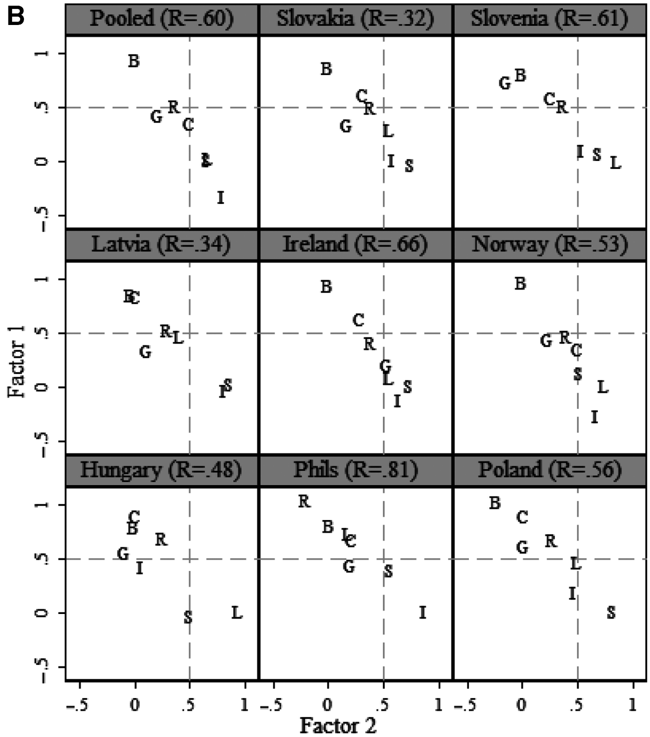
FIGURE 1 **By-country factor plots of Geomin Rotated Factor Scores from EFA for countries with well-fitting two-factor solution EFA two-factor solutions: (A) Group 1 and (B) Group 2.** *Note:* **B = birth, C = citizen, R = residence, G = religion, L = language, I = Institutions, S = sentiment. Items located above horizontal dashed line should be considered as influential for Factor 1 and those to the right of the vertical dashed line influential for Factor 2**



language is and how difficult the language is to learn. Residence, for example, may be more salient for ascriptive identity in countries that have strict entrance requirements. Table 3 shows that Born loads moderately to strongly on factor 1 in all countries and that Institutions and/or Sentiment loads moderately to strongly on factor 2 in all countries. Given these results, we are comfortable associating factor 1 with ascriptive identity and factor 2 with voluntarist identity. The remaining items vary in their level of association to each dimension. This, we believe, reflects the context-dependence of the concept of national identity and its link to these items.

To further validate that the LCFA models are meaningful, we examined three countries in more depth—the United States, Great Britain, and Canada—and contacted a regional expert familiar with national identity about two further cases—Slovenia and Slovakia. Figure 1A and B present scatter plots of the loadings of each indicator on factor 1 by their loading

Figure 1 **Continued**



on factor 2. Such plots allow a quick evaluation of differences and similarities of construct form across countries, as well as between each country and the pooled model. The dashed lines in each plot show which indicators are most influential for each factor—as a general rule, those above the horizontal line inform factor 1, whereas those to the right of the vertical line inform factor 2.

Given their intertwined political histories, one might expect LCFA models in the United States, Canada, and Great Britain to be similar. Indeed, Figure 1A indicates that there are similarities across these models. In particular, the models for the United States and Canada bear a striking resemblance to one another: in both countries Born, Religion, Residence and Citizenship inform factor 1, while Sentiment and Institutions inform factor 2. The most striking difference between the country plots is with the position of Language. In the United States, language is most closely associated with factor 1. In Canada, however, Language is not an important indicator for either factor. This likely reflects the diminished saliency of language as it relates to national identity in Canada, given their adoption of multiple national languages. This was reflected in the question itself, which in Canada asked about the importance of speaking either English or French.

Comparing Great Britain to either Canada or the United States reveals further differences. For one, in both the United States and Canada, Citizenship informs factor 1—linked with both Born and Residence. In Great Britain, however, these three indicators are separated, with Born informing factor 1, but Citizenship informing factor 2—and Residence proving unimportant to either. Britain's colonial history may help to explain these differences. For example, Citizenship is likely decoupled from Born as individuals born to British citizens living in British colonies—from India to South Africa to the Caribbean—prior to their respective independence were still considered British. As Residence crosscuts both Citizenship and Born, with many British citizens, whether born in Great Britain or not, living abroad in British or former British colonies, its ambiguous importance is historically reasonable. Figure 1A also shows how Language informs national identity differently in Great Britain compared to the United States and Canada—in the former, it loads on factor 2. This likely reflects the English language as an identifying characteristic of "being British" or having a British identity that transcends birthplace or residence.

We also confirmed the meaningfulness of LCFA models for Slovenia and Slovakia by consulting with a political sociologist with expertise on national identity in post-Soviet Europe. As surveys get exported from Western to non-Western countries, there are increasing concerns about how adequately items constructed from Western ideas translate to new countries. Accordingly, we picked two post-Soviet countries to assess whether the models given with LCFA could be evaluated as meaningful constructs of national identity within each country. Unlike our own evaluations, however, which relied on post-hoc explanations of differences between the constructs in Canada, the United States and Great Britain, for these evaluations we sent our expert a list of the indicators and asked her to hypothesize as to the relevant importance of each indicator in the concept of national identity. Our expert's comments were in line with the findings of the LCFA models. With regards to religion, for example, she hypothesized that it would be more important for Slovenia than Slovakia, and specifically that it would be associated with ascriptive national identity (factor 1). In explaining this distinction, she indicated that while religion under Communist rule in Slovakia was essentially eradicated, in Slovenia the Communist government made concessions to the Catholic Church during its rule and religion continued to prosper. Our expert also described religion as the "lifeblood" of Slovenian society. Indeed, the LCFA model for Slovenia shows religion loading very strongly on factor 1—alongside considerations of birth. Conversely for Slovakia, Religion fails to inform either factor strongly.

Language was another important difference between these two countries. Our expert indicated that given the similarities of language between Slovakia and the Czech Republic—Slovak and Czech are essentially indistinguishable—language should be a less important indicator in Slovakia than in Slovenia in regards to voluntarist identity. As Slovenian shares few linguistic characteristics with any other world language, however, she anticipated it would be of strong import for the voluntarist component of national identity. Both of these hypotheses were confirmed by the LCFA model results shown in Figure 1B.

Such evaluations of the LCFA models serve to highlight that while the concept being measured in each country is equivalent, the constructs representing the concept in each country are informed by specific context. Differences between the country-specific models—as well as with the pooled model—assure us that the LCFA approach can provide additional information not found within pooled models, and thus produce useful, country-specific (but also functionally equivalent) constructs.

Comparing Approaches

The by-country factor plots in Figure 1A and B illustrate that there are substantial differences between the constructs arrived at through the LCFA and the pooled EFA approach. If the assumptions of the pooled EFA held, we would expect these patterns to be similar. Table 5 presents the squared correlations between each measure of national identity, indicating the proportion of shared variance.

For factor 1, squared correlations range from a low of 0.79 between the LCFA and the summative scales in New Zealand to a high of 0.996 between the pooled EFA and LCFA scales in Norway. For factor 2, correlations range from a low of 0.49 between the LCFA and summative scales in Hungary to a high of 0.99 between the pooled EFA and LCFA scales in Great Britain. Generally, correlations are higher between the pooled EFA and LCFA scales. Additionally, correlations are generally higher for factor 1 scales than factor 2.

Next we examine if substantive conclusions vary by measurement approach. That is, if our intent is to compare national identity across any two countries, could our choice of measurement determine our findings? We do this in two ways. First, we examine, *within* any particular country, whether covariates for predicting national identity differ in significance by measurement model approach. Second, we examine whether differences in covariate effects (again in terms of significance) *between* any two countries vary by measurement model approach. Table 6 shows an example of both of these comparisons, for four countries and two covariates, for factor 2.

To evaluate within-country differences, we ran country-specific regression models on each measure of national identity, using the same set of independent variables. Table 6 shows extracts from the larger results—two covariates predicting factor 2 for the United States, Canada, Slovenia and Slovakia. For

TABLE 5 **Squared correlations between constructs for national identity, by factor, country, and method**

Correlation between	*Factor 1*			*Factor 2*		
	EFA & LCFA	*EFA & Sum*	*LCFA & Sum*	*EFA & LCFA*	*EFA & Sum*	*LCFA & Sum*
All countries	0.94	0.94	0.90	0.91	0.83	0.83
Australia	0.97	0.87	0.87	0.96	0.68	0.76
Great Britain	0.97	0.89	0.84	0.99	0.81	0.78
Netherlands	0.99	0.87	0.87	0.96	0.69	0.68
Sweden	0.98	0.88	0.91	0.92	0.65	0.70
U.S.	0.95	0.89	0.93	0.78	0.75	0.88
Austria	0.91	0.88	0.92	0.81	0.76	0.89
Hungary	0.84	0.85	0.93	0.64	0.63	0.49
New Zealand	0.99	0.85	0.79	0.96	0.66	0.71
Latvia	0.91	0.85	0.89	0.73	0.70	0.87
Slovakia	0.97	0.84	0.88	0.91	0.69	0.81
Slovenia	0.91	0.87	0.96	0.94	0.73	0.79
Ireland	0.97	0.87	0.87	0.96	0.80	0.73
Norway	1.00	0.88	0.86	0.98	0.66	0.66
Canada	0.97	0.90	0.93	0.58	0.70	0.70
Philippines	0.83	0.87	0.82	0.94	0.88	0.93
Poland	0.93	0.89	0.92	0.84	0.75	0.88

Note: Squared correlations indicate the proportion of shared variance.

example, within Slovenia parental citizenship is a significant predictor of voluntarist national identity under the pooled EFA and LCFA approaches, but not for the summative scale. In Canada, a preference for immigrant adaptation is significantly associated with voluntarist identity under the summative and pooled EFA approaches, but not under LCFA. For both countries, then, the significance of the variable depends upon the measurement model used. Table 7 summarizes the number of predictors whose significance varied by measurement model approach for each country. These differences range from a low of 8 percent over both factors for Australia, Sweden, and Slovakia—or one variable out of 12 (6 on each of 2 factors) varying by measurement model approach—to a high of 42 percent (5 of 12 variables) in New Zealand.

Next we turn to between-country differences. To test if the relationship between any covariate and national identity varies by country and by measure, we further estimate a fully interactive model, where each measure of national identity is regressed on each country dummy and each country-covariate interaction with the constant suppressed. A Wald test is used to test if two country-covariate interactions are equal (H_o: $\beta^{USA}_{adapt} = \beta^{Canada}_{adapt}$ for each measure of national identity and for each country-pair comparison). Table 6 presents results comparing four countries on two covariates for factor 2. Again, our interest is in significance that differs across approaches. Tests that differ

TABLE 6 **Examples of within- and between-country differences in effects by measurement model**

	Parent citizen			*Groups adapt*		
	Pooled	*Summative*	*LCFA*	*Pooled*	*Summative*	*LCFA*
USA	0.268* (0.125)	0.002 (0.127)	0.091 (0.129)	0.172*** (0.053)	0.218*** (0.054)	0.200*** (0.055)
Canada	0.191* (0.079)	−0.053 (0.081)	−0.107 (0.082)	0.216*** (0.051)	0.191*** (0.052)	0.064 (0.053)
Slovenia	0.317** (0.128)	0.199 (0.130)	0.302* (0.131)	0.073 (0.060)	0.065 (0.061)	0.075 (0.062)
Slovakia	0.449** (0.143)	0.551*** (0.145)	0.522*** (0.147)	0.296*** (0.053)	0.289*** (0.054)	0.310*** (0.054)
F test:						
$\beta_{USA} = \beta_{Canada}$	0.27	0.13	1.70	0.34	0.13	3.20
$\beta_{USA} = \beta_{Slovenia}$	0.08	1.18	1.31	1.51	3.46	2.26
$\beta_{USA} = \beta_{Slovakia}$	0.92	8.12**	4.88*	2.69	0.87	2.02
$\beta_{Canada} = \beta_{Slovenia}$	0.70	2.72	6.98**	3.24	2.44	0.02
$\beta_{Canada} = \beta_{Slovakia}$	2.51	13.24***	14.03***	1.18	1.72	10.57**
$\beta_{Slovenia} = \beta_{Slovakia}$	0.48	3.27	1.25	7.70**	7.53**	8.10**

Note: Effects of parental citizenship and feelings about immigrant adaptation on voluntarist national identity (Factor 2) for USA, Canada, Slovakia, and Slovenia. Standard errors are given in parentheses. The *F*-test indicates if the effects of the covariate are equal between countries. Shaded rows indicate that the equality of the covariate between countries differs by measurement model approach. Full results available at http://www.taitmedina.com/papers.html.

$^{*}p \leq .05$; $^{**}p \leq .01$; $^{***}p \leq .001$.

TABLE 7 **Summary of within-country effects**

	Factor 1 *N (%)*	*Factor 2* *N (%)*	*Total* *N (%)*
Australia	0 (0)	1 (17)	1 (8)
Sweden	0 (0)	1 (17)	1 (8)
Slovakia	0 (0)	1 (17)	1 (8)
Ireland	1 (17)	1 (17)	2 (17)
Great Britain	1 (17)	1 (17)	2 (17)
USA	1 (17)	1 (17)	2 (17)
Philippines	2 (33)	0 (0)	2 (17)
Poland	1 (17)	1 (17)	2 (17)
Canada	0 (0)	3 (0)	3 (25)
Netherlands	0 (0)	3 (50)	3 (25)
Austria	1 (17)	3 (50)	4 (33)
Hungary	1 (17)	3 (50)	4 (33)
Latvia	2 (33)	2 (33)	4 (33)
Slovenia	0 (0)	4 (67)	4 (33)
Norway	2 (33)	2 (33)	4 (33)
New Zealand	2 (33)	3 (50)	5 (42)

Note: Number and percentage (out of six) of covariate effects whose significance differs by measurement model.

TABLE 8 **Summary of between-country effects**

	Age (%)	*Higher ed (%)*	*Female (%)*	*Foreign (%)*	*Parent is citizen (%)*	*Adapt to culture (%)*
Factor 1: Ascriptive	18 (15)	28 (23)	16 (13)	15 (13)	15 (13)	11 (9)
Factor 2: Voluntarist	45 (38)	51 (43)	16 (13)	20 (17)	18 (15)	37 (31)

Note: Number and percentage (out of 120) of country-pair comparisons whose significance differs by measurement model.

across approaches are shaded in the table. For example, the effect of parent citizenship in Canada and Slovenia are equivalent under the pooled and summative approaches, but differ significantly under LCFA (F-test: Pooled: 0.70; Summative: 2.72; LCFA: 6.98).

Table 8 presents the percentage of country-pair comparisons out of a possible 120 comparisons that differ by measurement model for each covariate for each factor.

For factor 1, between 9 and 23 percent of conclusions vary by measurement approach. For factor 2, between 15 and 43 percent of conclusions are dependent on measurement approach.

These differential findings by measurement model indicate that *measurement matters and that results can vary by choice of method*. To the degree that such differences occur, we cannot rule out that it is measurement model choice that determines substantive conclusions.

DISCUSSION AND CONCLUSIONS

This article has examined four approaches to measurement modeling and shown that conclusions can be dependent on measurement model choice. Our goal in this article has been one of sensitization. While we cannot "prove" that one measurement model approach provides "truer" estimates of social relationships, the vast degree of difference across measurement models shown in the previous section strongly suggests that the measurement process *itself* should be a critical part of cross-national research and that analysts should be prepared to fully explain and defend their measurement decisions. A thorough understanding of the implicit assumptions of measurement modeling is required to avoid drawing conclusions that are little more than arbitrary.

Analysts should not expect cross-national measurement to be a simple application of boilerplate methods. Preeminent cross-national surveys have made it a priority to scrutinize indicators included in cross-national surveys to ensure their comparability. Analysts should be prepared to apply this same rigor to the construct level. It is a common misperception that scaling decisions do not matter—that all measurement models are simply variants of other models. As this article shows, in cross-national work the appropriateness of a measurement model depends upon assumptions about convergence or divergence in conceptual manifestation (*i.e.*, the construct) across countries. Summative scales, for example, might be an appropriate technique to employ if researchers are comfortable theorizing that all indicators inform their construct to an equivalent degree, and that this construct is equivalent across all countries. Other researchers, however, may find it substantively important to test whether these assumptions are supported, and as such turn to MGCFA. And still other researchers may have theoretical reasons to believe that conceptual manifestation differs by country, thus requiring different, yet functionally equivalent, constructs for each country. Just as survey developers have recognized that the literal translation of indicators does not always lead to maximum comparability, so too might analysts decide that the most meaningful, appropriate and valid constructs are functionally equivalent, not necessarily measurement invariant. In all three cases, however, analysts should look to theory to both guide and defend their measurement model decisions.

This article has also presented a new measurement model approach referred to as locally-conditioned factor analysis or LCFA. While we do not endorse LCFA as uniformly the "best" approach, we believe there are theoretical reasons for further investigations of this approach. Methods privileging measurement invariance are currently considered by many to be a ‘‘gold standard’’ in comparative research (*cf.* De Beuckelaer, Lievens & Swinnen, 2007); however if requirements of measurement invariance are instituted without theoretical foundation, they may serve to artificially restrict comparative analyses. From our perspective, measurement non-invariance is not a show-stopper, but rather an outcome to be explained. Non-invariance provides analysts with an opportunity to more closely consider sources of variation and how such variation maps onto measurement—and through such explorations come conceptual and theoretical development. There is opportunity, for example, for analysts to use macro-level data, such as country-level economic, political or cultural indicators, to explain patterns of variation across country-specific measurement models.

We do not see this article as a final step. Work remains in testing whether different measurement modeling approaches—and specifically LCFA—can provide analysts with more purchase in cross-national comparison. Simulation studies, for example, where the structure of the data can be specified, would allow investigation of whether incorrect measurement model choice can bias substantive results and conclusions. More importantly however, applying theoretically informed measurement models to well-established fields of knowledge may reveal relationships that have been artifacts of incorrect measurement model specification.

Cross-national measurement is difficult. The rapidly increasing availability of high-quality cross-national survey data has only served to raise the stakes for cross-national research. The ease with which we can compare a diverse set of countries on similar indicators is only beneficial if comparisons are meaningful. Our results highlight that measurement modeling undertaken without a view towards implicit assumptions leads to results that may be measurement-dependent.

REFERENCES

Almond, G. A., & Verba, S. (1963). *The civic culture: Political attitudes and democracy in five nations*. Princeton, NJ: Princeton University Press.

Anderson, B. (1983). *Imagined communities: Reflections on the origin & spread of nationalism*. London: Verso.

Batalova, J. A., & Cohen, P. N. (2002). Premarital cohabitation and housework: couples in cross-national perspective. *Journal of Marriage and Family, 64*, 743–755.

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley Series in Probability and Mathematical Statistics. New York: Wiley.

Bollen, K. A., & Long, J. S. (Eds.) (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Bulmer, M. (2001). Social measurement: what stands in its way? *Social Research, 68*, 455–480.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

Browne, M. W., & Curdeck, R. (1993). Alternative ways of assessing model fit. In: K. A. Bollen & J. S. Long (Eds.), *Testing structural equation modeling* (pp. 136–162). Newbury Park: Sage Publications.

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*, 456–466.

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1–27.

Duncan, O. D. (1984). *Notes on social measurement, historical and critical*. New York: Russell Sage Foundation.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*, 466–491.

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling, 13*, 378–402.

Fuwa, M. (2004). Macro-level gender inequality and the division of household labor in 22 countries. *American Sociological Review, 69*, 751–767.

Gesis (n.d.). *Background variables and further coding standards*. Retrieved June 3, 2009, from http://www.gesis.org/en/services/data/survey-data/issp/issp-members-area/coding-standards/.

Harkness, J. A., Mohler, P. P., & Van de Vijver, F. J. R. (2003). Comparative research. In: J. A. Harkness, F. J. R. van de Vijver & P. P. Mohler (Eds.), *Cross cultural survey methods* (pp. 3–16). New Jersey: Wiley.

Heath, A. F., Fisher, S. D., & Smith, S. N. (2005). Globalization of public opinion research. *Annual Review of Political Science, 8*, 49–71.

Johnson, T. P. (1998). Approaches to equivalence in cross-cultural and cross-national survey research. *ZUMA-Nachrichten Spezial, January*, 1–40.

Jones, F. L., & Smith, P. (2001). Diversity and commonality in national identity: an exploratory analysis of cross-national patterns. *Journal of Sociology, 37*, 45–63.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36*, 409–426.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavior Research, 39*, 479–515.

Muthén, B. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.

Muthén, B., & Christoffersson, A. (1981). Simultaneous factor analysis of dichotomous variables in several groups. *Psychometrika, 46*, 407–419.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–189.

Muthén, B., & Kaplan, D. (1992). A comparison of some methodologies for the factor analysis of non-normal Likert variables: A note on the size of the model. *British Journal of Mathematical and Statistical Psychology, 45*, 19–30.

Muthén, B., du Toit, S. H. C., & Spisic, D. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Accepted for publication in Psychometrika*. Available at: http://www.gseis.ucla.edu/faculty/muthen/articles/Article_075.pdf.

Muthén, B., & Asparouhov, T. (2002). Latent variable analysis with categorical outcomes: multiple-group and growth modeling in mplus. *Mplus Web Notes: No. 4*. Available at: http://www.statmodel.com/download/webnotes/CatMGLong.pdf.

Park, A., & Jowell, R. (1997), Consistencies and differences in a cross-national survey. *The International Social Survey Programme (1995)*. Available at: http://www.za.uni-koeln.de/data/en/issp/codebooks/ZA2880_mr.pdf.

Pescosolido, B. A., & Olafsdottir, S. (2008, June). *The logistics of survey implementation in a comparative study of mental illness: Issues and resolutions in translation across cultural boundaries*. Paper presented at International Conference on Survey Methods in Multinational, Multiregional, and Multicultural Contexts. Berlin, Germany.

Przeworski, A., & Teune, H. (1966). Equivalence in cross-national research. *The Public Opinion Quarterly, 30*, 551–568.

Przeworski, A., & Teune, H. (1970). *The logic of comparative social inquiry*. New York: Wiley-Interscience, John Wiley & Sons.

Semyonov, M., Raijman, R., & Gorodzeisky, A. (2006). Rise of anti-foreigner sentiment in european societies, 1988-2000. *American Sociological Review, 71*, 426–449.

Smelser, N. (1976). *Comparative methods in the social sciences*. Englewood Cliffs, New Jersey: Prentice Hall Inc.

Smith, A. D. (1991). *National Identity*. London: Penguin.

Smith, T. W. (2003). Developing comparable questions in cross-national surveys. In: J. A. Harkness, F. J. R. van de Vijver & P. P. Mohler (Eds.), *Cross cultural survey methods* (pp. 69–91). New Jersey: Wiley.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*, 78–90.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Thurstone, L. L. (1947). *Multiple-Factor Analysis*. Chicago: University of Chicago Press.

Van de Vijver, F. J. R., & Poortinga, Y. H. (1982). Cross-cultural generalizations and universality. *Journal of Cross-Cultural Psychology, 13*, 387–408.

Verba, S. (1969). The uses of survey research in the study of comparative politics: Issues and strategies. In: S. Rokkan, S. Verba, J. Viet & E. Almasy (Eds.), *Comparative survey analysis* (pp. 56–106). The Hague: Mouton & Co.

Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes (Doctoral dissertation, University of California, Los Angeles, 2002). Retrieved March 11, 2008 from http://www.statmodel.com/download/Yudissertation.pdf.

Zentralarchiv fuer Empirische Sozialforschung. ISSP 1995: National Identity I [Data file]. Koeln, Germany: Zentralarchiv fuer Empirische Sozialforschung/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors]. Retrieved November 9, 2006 from http://www.icpsr.umich.edu/.

BIOGRAPHICAL NOTES

Tait R. Medina is a doctoral candidate in Sociology at Indiana University-Bloomington. Her research focuses on the impact of social forces on health, illness and healing, statistical methods, and institutions. Currently she is serving as research analyst on the Stigma in Global Context-Mental Health study, a National Institute of Health-funded project designed to study the etiology of the stigma of mental illness across 15 countries.

Shawna N. Smith is a doctoral candidate in Sociology and Jacob K. Javits fellow at Indiana University-Bloomington. Her research focuses on cross-national policy and law, institutions, regulation and governance, and statistical methods. Her dissertation examines the trajectories of policy related to equal employment in the US, Canada and the UK. Prior to IU, she received her MPhil in Sociology at Oxford University as a US-UK Fulbright Exchange Scholar.

J. Scott Long, Chancellor's Professor of Sociology and Statistics at Indiana University-Bloomington and Associate Vice Provost for research. His research focuses on gender differences in the scientific career, stigma and mental health, aging and labor force participation, human sexuality, and statistical methods. His research on the scientific career was published as From Scarcity to Visibility published by the National Academy of Sciences. He is past Editor of Sociological Methods and Research and the recipient of the American Sociological Associations Paul F. Lazarsfeld Memorial Award for Distinguished Contributions in the Field of Sociological Methodology. He is author of Confirmatory Factor Analysis, Covariance Structure Analysis, Regression Models for Categorical and Limited Dependent Variables, and Regression Models for Categorical and Limited Dependent Variables with Stata (with Jeremy Freese) and several edited volumes. His most recent book in 2008 is The Workflow of Data Analysis Using Stata.

Address correspondence to Tait R. Medina, Department of Sociology, Ballantine Hall 744, Indiana University, 1020 E. Kirkwood Avenue, Bloomington, IN 47405, USA, E-mail: tmedina@indiana.edu