



World Values Research WVR

Volume 8 / Number 5 / 2015

Misconceptions of Measurement Equivalence: Time for a Paradigm Shift

CHRISTIAN WELZEL *Leuphana University Lüneburg**
RONALD INGLEHART *University of Michigan at Ann Arbor*

ABSTRACT

Structural equation modelers judge multi-item constructs against three requirements: (1) multiple items converge in a single dimension; (2) individual-level patterns of item convergence are invariant across countries; (3) aggregate-level patterns of item convergence replicate those at the individual level. This approach involves two premises: measurement validity hinges solely on a construct's internal convergence and convergence patterns at the individual level have priority over those at the aggregate level. We question both premises (a) because convergence patterns at the aggregate-level exist in their own right and (b) because only a construct's external linkages reveal its reality outreach. In support of these claims, we use the example of "emancipative values" to show that constructs can entirely lack convergence at the individual level and nevertheless exhibit powerful and important linkages at the aggregate level. Consequently, we advocate a paradigm shift from internal convergence towards external linkage as the prime criterion of validity.

Key words: equivalence, external linkage, internal convergence, validity.

Word count: 7,216 (version accepted October 1st 2015).

CITATION

Welzel, C. & R. Inglehart (2015). "Misconceptions of Measurement Equivalence: Time for a Paradigm Shift." *World Values Research* 8(5):151-173.

* Christian Welzel is chair of Political Culture Research at Leuphana University Lüneburg, Visiting Professor at the Higher School of Economics, St. Petersburg, Russia, and Vice-President of the World Values Survey Association. Ronald Inglehart is a Distinguished Professor at the University of Michigan at Ann Arbor, Visiting Professor at the Higher School of Economics, St. Petersburg, Russia, and Founding President of the World Values Survey Association. Send correspondence to cwelzel@gmail.com.

Misconceptions of Measurement Equivalence: Time for a Paradigm Shift

INTRODUCTION

In a recent article, Aléman/Woods (2015) question well-known measures of human values from the World Values Surveys (WVS). Specifically, the authors criticize Inglehart/Welzel's (2005: 48-56) "traditional vs. secular-rational values" and "survival vs. self-expression values" as well as the updates of these constructs labeled "secular values" and "emancipative values" by Welzel (2013: 59-73).

All four of these constructs are multi-item combinations. They have been used in dozens of publications¹ for a number of meaningful purposes, including: (a) mapping cultural differences across the major nations of the globe; (b) tracing change in human values over time and across generations; (c) linking trajectories of cultural change to basic socio-economic transformations, including rising life expectancies, expanding literacy, education and informational connectedness, advancing globalization and urbanization, technological progress and improving living standards; (d) explaining differences in national policy outcomes, from gender equality to sexual liberation to peace and sustainability; and (e) modelling regime dynamics, such as where democracy emerges and flourishes and where impartial governance and other aspects of institutional quality prevail and where all of this is not the case.²

Aléman/Woods raise doubts against this line of research by the sweeping claim that none of the four involved value constructs is valid. The authors derive this postulate from a simple finding that they take a long way to demonstrate: the individual-level coherence of the items that constitute the four value constructs varies across countries and is weak in many of them.³ Aléman/Woods conclude from this presumably groundbreaking discovery that all four value constructs are incomparable across cultures. Any cross-cultural pattern found with these constructs, such as the Inglehart-Welzel cultural map of the world, is hence unreal. The obvious implication is to abandon these constructs altogether.

¹ A selection of the most important works includes Inglehart 1997; Inglehart/Norris/Welzel 2003; Inglehart/Norris 2003; Welzel/Inglehart/Klingemann 2003; Inglehart/Welzel 2003; 2005; 2010; Norris/Inglehart 2004; Welzel/Inglehart/Deutsch 2005; Welzel/Inglehart 2010; Welzel/Deutsch 2012; Welzel 2006; 2007; 2013; Welzel/Dalton 2014.

² Some findings in this body of work have initiated debates. For instance, Spaiser et al. (2014), Spaiser/Sumpter (2015) and Dahlum/Knutzen (2015) question the impact of emancipative values on the emergence and florescence of democracy. In response to this criticism, Welzel/Inglehart/Kruse (2015) demonstrate conclusive evidence for precisely that impact.

³ Aléman/Woods' apply multi-group confirmatory factor analysis to show that there is "configurational," "metric" and "scalar" variance in how the constituent items of our value constructs relate to each other in each country. Basically, this is about cross-country variability in inter-item coherence at the individual level.

Aléman/Woods' critique resonates with a growing consensus on the meaning of measurement equivalence. In a nutshell, this consensus centers on the following axiom: when one uses individual-level data to create a multi-item index and then calculates country-level averages on this index, these averages are equivalent only if they represent in each country the same individual-level pattern of inter-item coherence (cf. Stegmüller 2011; van de Vijver 2011; Davidov et al. 2012). As we will argue, this notion of equivalence is fundamentally flawed and needs to be replaced by a decidedly "nomological" view: numerically similar scores of a construct are equivalent whenever they result from similar antecedents or lead to similar outcomes at their level of analysis.

We do not debate Aléman/Woods' evidence that our value constructs show variable and often weak inter-item coherence at the *individual level within countries*. In fact, we ourselves have pointed out this phenomenon very explicitly a number of times (Inglehart/Welzel 2003; 2005: 231-244; Welzel 2013: 74-79, 110-112).

We do, however, reject Aléman/Woods' conclusion that a construct's country-level scores are incomparable merely because they emerge from variable individual-level coherence patterns. Our rejection is informed by a proper understanding of the "micro-macro puzzle," which allows for only one conclusion in this matter: a construct's measurement features at the individual level provide *no information whatsoever* about the same construct's validity at the country level. As we will outline, this insight rectifies a widespread misunderstanding of the "ecological fallacy."

Our second contribution is to explain why inter-item coherence is a misleading standard that has created a false sense of "measurement equivalence"—a false sense especially when one evaluates constructs based on a *combinatory* instead of a *dimensional* logic. Many scholars are not aware of this distinction, which is known in the methods literature as the difference between "formative" and "reflective" constructs (Coltman et al. 2008). It is, hence, worthwhile to explain this distinction and to point out its implications for measurement equivalence.

Our third contribution is to demonstrate that cross-national differences in a construct's inter-item coherence are *not* an infallible sign of in-comparability. On the contrary, if these differences map closely on a "third" standard of reference, then these differences embody a common meaning—the very basis of comparability. We will evidence this point for the construct of emancipative values: the individual-level coherence among this construct's items increases systematically with a major aspect of modernization at the country level: cognitive mobilization. As a construct designed to capture key features of modernization, emancipative values are not invalidated by the fact that certain properties of their measurement, including inter-item coherence, are themselves shaped by modernization. The exact opposite is the case.

In conclusion, we maintain that our value constructs are as valid, meaningful and real as most of the literature treats them. We nevertheless appreciate Aléman/Woods' criticism because it provides an overdue opportunity to highlight some widespread misconceptions in the assessment of measurement equivalence.

UNDERSTANDING THE MICRO-MACRO PARADOX

Although the data we use to create our value constructs are collected from individuals, the purpose of these constructs is not to measure internally coherent personality traits. Item-response coherence inside individuals is beyond our intention. Instead, we aim to capture patterns in value orientations that emerge first and foremost--and sometimes solely--at the group level, the place where culture takes shape. Value patterns at the group level, most notably countries, describe prevalence features in collective mentalities. Since Aléman/Woods evaluate our constructs exclusively at the individual level, they miss the whole point of our intention, even though we stressed from the beginning that the individual level is not the place where the patterns of our interest exist. Some of the most pronounced value patterns are simply invisible at the individual level, for they represent a culture-type phenomenon that only surfaces in the aggregate.

Against this backdrop, it is helpful to remember what Kaase (1986) termed the "micro-macro-puzzle in the social sciences." Referring to Converse's (1964) discovery of "ideological incoherence" among individuals, the micro-macro puzzle denotes the well-known fact that incoherence between attitudes at the individual level contrasts sharply with coherence among aggregate measures of the same attitudes. With respect to public opinion, which is "public" only in the aggregate, Page/Shapiro (1992) and Stimson/McKuen/Erikson (2002) provide ample evidence for this pattern. It is visible in the fact that a given pair of attitudes regularly correlates much more weakly at the individual level within countries than aggregate measures of the same two attitudes correlate across countries.

For instance, the "choice" and "voice" components of emancipative values⁴ correlate at $R = 0.20$ at the individual level, in contrast to an R of 0.68 at the aggregate level.⁵ With regard to multi-item constructs, the micro-macro puzzle means that these constructs always reach stronger coherence at the aggregate than at the individual level.⁶ An obvious conclusion from this observation is that weak inter-item coherence at the individual level is (a) the norm and (b) irrelevant for coherence patterns in the aggregate.

⁴ Emancipative values cover twelve WVS-items that tap in various ways an emphasis on freedom of choice and equality of opportunities. The items group into four sub-indices (each consisting of three items), indicating an emphasis on (1) people's "voice," (2) reproductive "choice," (3) child "autonomy" and (4) gender "equality." Each of the four sub-indices is standardized on a scale with minimum 0 and maximum 1. The overall index of emancipative values measure (EVI) is the average across these four sub-indices. Theoretical background, measurement procedure as well as reliability and validity statistics are detailed in Welzel (2013: 57-103).

⁵ Data cover 58 nations and 78,000 individuals from the sixth wave of the WVS conducted in 2012-14. Individual- and aggregate-level correlations are similar for other waves of the WVS.

⁶ The opposite case—stronger coherence at the individual than at the aggregate level—is a logical possibility, of which we know not a single example.

Inglehart/Welzel (2003; 2005: 231 - 244) propose three causes of the micro-macro puzzle. First, individual-level survey data are inflicted with large amounts of measurement error, known as "background noise," that partially obscures true correlations. However, much of the error is distributed randomly, so aggregation cancels most of it out--reducing a source of pattern disguise.⁷ Second, pooling aggregated data across diverse sets of countries widens variability in correlated variables beyond the "clouds" within which correlations remain murky (and much of the within-country variation is inside the clouds⁸). Third, countries are powerful entities of socialization whose cultural reproduction over subsequent generations creates inert trajectories: this implies pattern consistency in the aggregate. Culturally ingrained orientations are no exception: such important orientations as out-group trust and racial tolerance cohere much more closely at the aggregate than at the individual level.⁹

It follows from all of this that the individual level of analysis is unsuited to reveal some of the most striking coherence patterns in orientations. Many coherence patterns—in particular those defining culture—only exist in the aggregate.

Moreover, the same attitudes mean something different at different levels of analysis (Leung/Bond 2007). Scores in emancipative values at the individual level indicate how much a given person prioritizes this set of values. At the aggregate level, the average in emancipative values indicates how prevalent these values are in a given country.¹⁰ In other words, the same variable measures personal preference strength at the individual level but social norm strength at the aggregate level. These are two different things, of which only the latter says something about culture. Accordingly, the correlation among the same pair of attitudes also captures something different at the individual and aggregate levels.

⁷ The sources of individual-level measurement error are manifold. For instance, respondents might give random answers either because they do not understand questions or because they have no solid preferences. Likewise, response styles--such as the urge to avoid extreme answers (moderateness)--might affect answers. Furthermore, there might be fraudulently duplicated responses. But insofar as these invalid answers distribute randomly, they leave aggregate patterns--such as marginal distributions and correlations--unaffected. We analyzed the entire WVS data for these three sources of measurement error and we found evidence for all of them: to some extent there are contradictory, moderateness-induced and duplicate responses. Yet, in the overwhelming majority of samples the extension of these problems is minor and rarely affects marginal distributions and correlations. Results are available upon request from us.

⁸ The stronger distributions on given variables cluster on country means, the more of the within-country variation is "clouded."

⁹ Based on WVS waves five and six, the average individual-level correlation within countries between out-group trust and racial tolerance is $R = 0.14$ ($N = 58,530$ individuals). The cross-country correlation between aggregate measures of the same two attitudes is $R = 0.50$ ($N = 75$ countries).

¹⁰ This statement assumes that individual-level scores in emancipative values are mean-centered and single-peaked in most countries. Welzel (2013: 84-86) shows that this assumption holds true.

To illustrate this point, let us consider the correlation between out-group trust¹¹ and emancipative values. At the individual level within countries, the correlation between these two orientations tells us to what extent a person who trusts more than most others in her country is also more emancipatory in her orientations than most others in this country. The correlation between country-level aggregations of the same two variables tells us something else: to what extent the prevalence of trust associates with the prevalence of emancipatory orientations. The existence of this association in its given strength is in no plausible way invalidated by the absence of a similar association at the individual level.¹² Patterns at different levels of analysis simply have no inferential value to each other, for they mean separate things.

THE NATURE OF ECOLOGICAL EFFECTS

For all of the above reasons, it can happen that strong coherence among aggregate measures of a set of attitudes coexists with variable and often weak coherence among the same attitudes at the individual level. Actually, the social nature of human existence makes this pattern quite common—through “ecological” effects. Ecological effects are manifestations of social influence. They operate in such a way that a given orientation shapes people’s other psychological and behavioral traits through the social prevalence of the orientation in question, irrespective of whether or not a person herself embraces the respective orientation. This is a very common regularity that Welzel (2013: 110-112, Box 3.1) has theorized as an “elevator phenomenon.”

An example is the association between emancipative values and nonviolent protests. The association exists at the individual-level within countries: persons with stronger emancipative values tend to participate in nonviolent protests more frequently. However, the correlation is only modestly strong and another pattern in the association between emancipative values and nonviolent protests is much more powerful¹³: when emancipative values become more prevalent in a country, everyone’s protest activity “elevates,” regardless of whether the person in question herself endorses emancipative values (Welzel 2013: 230).

Elevator effects of this kind are real and they shape societies powerfully, by influencing the prevalence of psychological and behavioral traits. But elevator effects are inherently ecological in character and, hence, not mirrored in a corresponding individual-level association. Accordingly, there exist meaningful and consequential coherence pat-

¹¹ We use the out-group trust measure introduced by Delhey/Newton/Welzel (2012).

¹² The cross-country correlation between aggregate measures of these two variables is $R = 0.68$ ($N = 75$ countries). The average individual-level correlation within countries between the same two variables is $R = 0.24$ ($N = 64,317$ individuals). Data are taken from WVS waves five and six.

¹³ Based on WVS waves one to five, the average individual-level correlation within countries between emancipative values and protest activity is $R = 0.37$ ($N = 301,808$ individuals), while the cross-country correlation between aggregate measures of these two variables is $R = 0.73$ ($N = 91$ countries).

terms in the aggregate, which have no equivalent at the individual level. But the latter does not render unreal the former.

In light of this, Przeworski/Teune's (1970: 73) famous dictum that "an ecological correlation is spurious, if it is not reflected in the same way at the individual level within each aggregate unit," is profoundly flawed. Scholars continue to recite this quote as a warning against the "ecological fallacy." But the irony is that this very statement is itself a flagrant illustration of a fallacy in the opposite direction of inference, known as the "individualistic fallacy": inferring the validity of an aggregate-level relation from its presence or absence at the individual level.

The ecological principle that shapes much of social reality also shapes multi-item composite measures of values, including emancipative values: because of ecological effects, the inter-item coherence of these values powerfully surfaces at the aggregate level between countries while remaining variable and weak at the individual level within countries (Welzel 2013: 74 - 79). Accordingly, one cannot assess the equivalence and other qualities of an aggregate measure, such as country averages in emancipative values, at the individual level—even if the aggregation derives from individual-level data.

Yet, judging the equivalence of country-level averages from consistency patterns at the individual level within countries is exactly what multi-group confirmatory factor analysis does—the new booming industry in cross-national survey research (cf. Stegmueller 2011; Davidov et al. 2012).¹⁴ No doubt, multi-group confirmatory factor analysis (MGCFA) is an excellent tool to examine item sets for dimensional unity at different levels of analyses. But dimensional unity is no criterion for combinatory constructs, which explicitly allow for multi-dimensionality.

DIMENSIONAL AND COMBINATORY LOGICS

MGCFA follows a "latent variable" logic that dominates the field of structural equation modelling (SEM). From the viewpoint of SEM, multi-item constructs are justified if—and only if—all included items converge in a single dimension and show no variability in this feature across countries. This approach starts from two premises that its adherents take for granted when in fact they are questionable upon closer scrutiny.

¹⁴ An example discussed in Inglehart/Welzel (2005: 232-233) is the relationship between unemployment and the Nazi vote in late Weimar Germany. At the individual level, unemployed people might not have been more prone to vote the Nazis than employed ones—in which case there was no individual-level linkage. At the aggregate level, however, regions with higher unemployment definitely voted in greater shares for the Nazis. The likely reason is that everyone in these regions became scared by unemployment, no matter of whether the person in question herself was unemployed or not. Would we follow Przeworski/Teune and other alerters of "aggregation bias," the existing cross-regional relationship between unemployment and the Nazi vote needed to be declared unreal if there was no corresponding individual-level relationship between unemployment and the Nazi vote. The absurdity of this conclusion should be obvious.

The first premise is that multi-item constructs only make sense when their constituent items approximate the condition of representing inseparable manifestations of a single dimension. If this requirement is met, it is justified to extract from the constituent items a summary measure, which "reflects" their unifying dimension. In this logic, divergent variance among constituent items is simply measurement error. Consequently, validity hinges entirely on item convergence. Constructs whose constituents do not strongly converge are "unreal" in this view.

We agree that the dimensional logic is a useful guide of construct formation for some purposes. But there is a powerful alternative logic that informs many of the most well-known multi-dimensional constructs. These constructs can be as meaningful as one-dimensional ones and are often more consequential than those (see footnote 20 for an explanation why). Alexander/Welzel's (2011a) two-dimensional construct of "effective democracy" is a case in point.¹⁵ Effective democracy consists of "democratic rights" as the base component and "law enforcement" as the factor that makes the base effective. The two constituents are not supposed to be convergent. On the contrary, they are combined because they tap distinct properties of effective democracy, which is theoretically predefined as the interaction of these properties.¹⁶ Hence, to measure effective democracy in its predefined meaning one must measure the combined presence of these properties—no matter how strongly they correlate. Contrary to dimensional logic, this combinatory logic requires constituent components to be recognizably divergent¹⁷, not convergent. The reason is that a combination of components can only make a difference, relative to what each single component does, when these components cover separate things. Their combination is justified merely by the fact that they represent mutually complementary qualities under an over-arching idea.

In the combinatory logic, divergent variance among constituent components is not considered as measurement error but as complementary reality coverage. The methods literature characterizes combinatory constructs as "formative" and juxtaposes them to the dimensional logic of "reflective" constructs (Coltman et al. 2008). Goertz (2006: 10-11) addresses the same distinction by the terms "latent" versus "ontological" constructs.

At any rate, it should be clear that item convergence is an inadequate criterion when the logic of construct formation is combinatory. This is important to note because

¹⁵ Against criticism by Hadenius/Teorell (2005) and Knutsen (2010), Alexander/Inglehart/Welzel (2012) demonstrate the validity of the effective democracy index (EDI) in comparison to the six most widely used measures of democracy, including Freedom House and Polity IV.

¹⁶ Because the definition focuses on an interactive combination, the two components are combined multiplicatively. With multi-dimensional constructs, multiplication is appropriate when the constituent components are supposed to condition each other. When, by contrast, the components are supposed to represent complementary qualities, an additive combination is appropriate (Welzel 2013: 259, Box 8.1).

¹⁷ "Recognizably different" does not require entirely different (i.e., no zero-correlation requirement). In combinatory logic, there is no precise requirement as exactly how different constituent components need to be; the only requirement is that the difference is significant.

Welzel's (2013: 60) measure of emancipative values is introduced explicitly as a combinatory construct, not a dimensional one. Specifically, the twelve items over which emancipative values are measured are portrayed as additive qualities under the definition of emancipation. Thus, to measure a subject's overall response to matters of emancipation, partial responses must be added up over all relevant items, in deliberate disregard of how consistent the partial responses appear throughout the item set. To measure an overall response on a defined field, such as emancipation, consistency among the partial responses is simply no requirement.¹⁸ Accordingly, the combinatory logic assumes mutual substitutability among partial responses (Goertz 2006: 10-13). And substitutability is an accurate assumption when variability in the composition of partial responses does not affect how an overall response relates to its expected antecedents and consequences.

The quality criteria for combinatory constructs are two-fold. Theoretically, the combination must make sense such that the components meaningfully complement each other under an over-arching definition. Empirically, the combination must make a difference in that it maps better on its expected antecedents or consequences than does each of its components. Consequently, the yardstick to judge a combinatory construct is its predictedness and predictiveness relative to other aspects of reality. Datler/Jagodzinski/Schmitt (2013) call this criterion external validity¹⁹, in contrast to internal consistency. From an epistemological point of view, external validity is the preferable criterion²⁰: when we have predictive power, we also have interventionist potential to change things in a desirable direction—the ultimate purpose of science.

If Aléman/Woods had assessed the value constructs of Inglehart/Welzel under external validity, their conclusions had to be radically altered. Inglehart/Welzel and their co-authors have shown in scores of publications that their value constructs associate at exceptional strength with several dozen key indicators of (1) socio-economic development, (2) cultural legacies and (3) institutional performance—which are some of the most fundamental aspects of societal existence (cf. Inglehart/Welzel 2010). The correspondence of the value constructs with these aspects of social reality ranges from sixty to eighty percent, across almost a hundred countries representing more than ninety percent of the world population. Whatever the causality behind these associations might be, they are so pervasive that there can be only one conclusion: these value constructs tap something real.

¹⁸ The exception is when components are strongly negatively correlated. Then their combined measure would not make sense because they cancel each other out (as strength in one component coincides with weakness in the other).

¹⁹ An alternative term is "nomological" validity: a construct should be associated reasonably strong with its theoretically expected correlates, be they antecedents, concomitants or consequences of the construct in question (cf. Denton 2008; Jackman 2008).

²⁰ It is indeed a matter of priority because external validity and internal consistency are in a trade-off relation, which is an aspect of the "bandwidth-fidelity dilemma": when a construct covers a wider scope of reality, this wider scope will make it indicative of more things and, hence, give it more explanatory power. At the same time, this wider scope makes the construct more diverse, which will decrease the coherence among its components (cf. Ones & Viswesvaran 1996).

The same is true of the widely cited Inglehart-Welzel cultural map. The pattern behind this map is so robust in the aggregate that it re-occurred in almost identical shape throughout six consecutive waves of the WVS, despite the fact that the country composition has been considerably changing from wave to wave. Moreover, the two dimensions on this map correlate strongly with other measures of cultural differences, taken from different data under the guidance of different concepts. For instance, the constructs of "individualism/collectivism" and "autonomy/embeddedness" share almost eighty percent variation with Inglehart/Welzel's (2005: 137) value constructs. Strong correlations also exist with a society's geo-climatic, genetic, nutritional and linguistic features—all of which testify to the validity of the value constructs from the WVS (Welzel 2013: 122).²¹

MISJUDGMENTS OF MEASUREMENT EQUIVALENCE

The second questionable premise is that cross-national variability in a construct's item convergence is an infallible indication of incomparability (cf. Stegmueller 2011; van de Vijver 2011; Davidov et al. 2012; van Deth 2013). From this premise one had to conclude that the same overall scores in emancipative values mean something different when they emerge from different compositions of partial scores.

Adopting this conclusion would overlook the possibility of compositional substitutability (Goertz 2006: 10-11): the same overall performances across a thematic field might emerge from different compositions of partial performances but nevertheless lead to similar consequences (or result from similar antecedents). Whenever this pattern exists, it is the overall performance over the field, not the composition of its partial performances, that matters—a clear case of compositional substitutability. Of course, the theoretical challenge is to identify thematic fields of such obvious relevance.

Let us consider the field of emancipative values under these auspices. Welzel (2013: 84-86) shows that individual-level distributions over the item set of emancipative values are strongly mean-centered and single-peaked in each country, giving the term "central tendency" real meaning. Welzel (2013: 74-79) also demonstrates that the constituent components of emancipative values cohere in widely different strength in different countries. But he does not conclude from this finding that the same country scores in emancipative values are incomparable.

The reason to not jump to this conclusion is straightforward: two numerically similar scores in a given measure are comparable, if their very similarity maps in corre-

²¹ To be precise, these correlations prove "nomological" validity, also known as "convergent" validity. In contrast to the latter, there is also the notion of "discriminant" validity: a construct should not be correlated to phenomena to which there is no theoretical connection (Jackman 2008). Emancipative values meet this criterion too. For instance, there is no clear-cut theoretical connection of emancipative values to "institutional trust" or "norm conformity." Accordingly, emancipative values do not correlate with these phenomena, neither at the individual nor the aggregate level (results available upon request).

sponding fashion on a “third” standard of reference, a so called tertium comparationis. Thus, comparability properly understood boils down to external validity, not internal consistency. External validity in turn is a matter of a construct’s association with its theoretically expected correlates, whether these correlates operate as antecedents, consequences or concomitants of the construct in question. In other words, the strength of a construct’s associations with its supposed correlates reveals how well this construct maps on “third” standards of reference. External linkage, in this sense, is actually the foremost measure of a construct’s reality coverage.

COMPARABILITY TESTING BEYOND COHERENCE

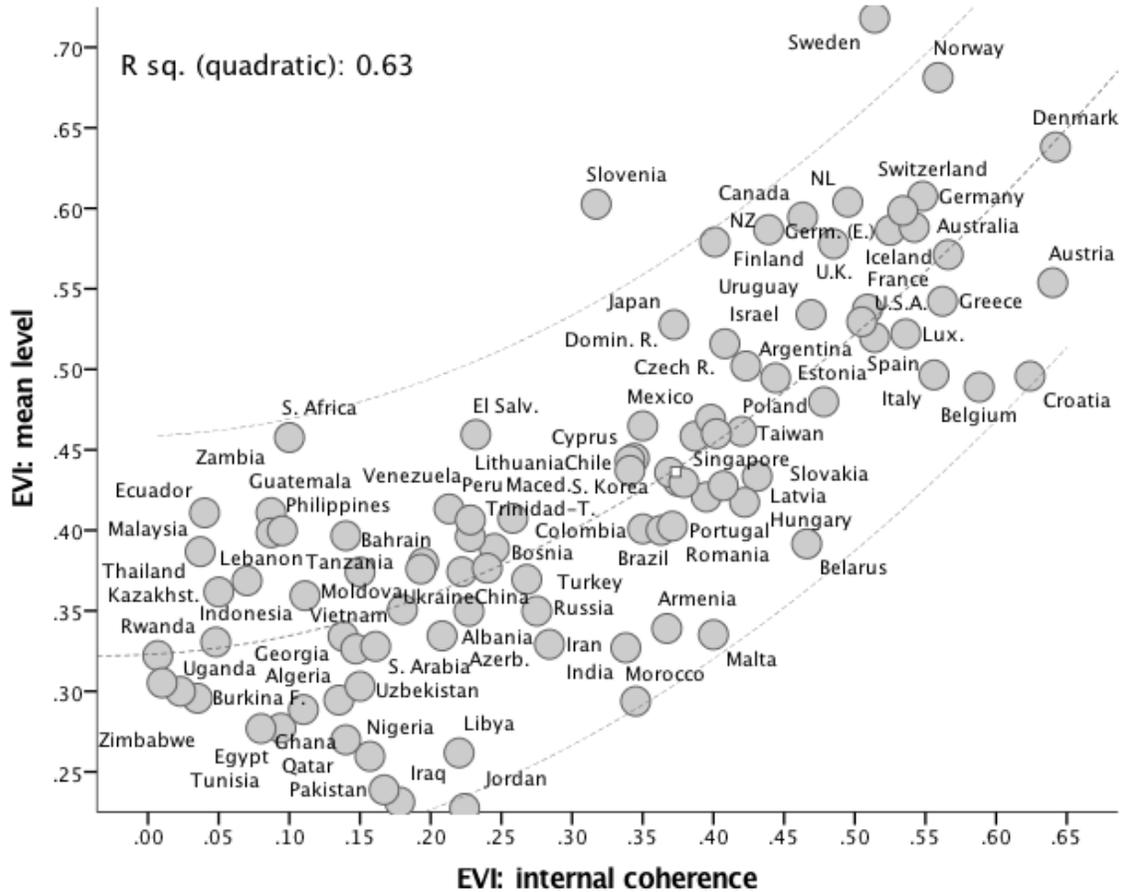
Cross-national variability in item convergence is too premature a finding to jump to the conclusion of incomparability. To come to conclusive inferences concerning this issue, three questions need close examination.

The first question is how much variability in the strength of item convergence exists independent of the country means in emancipative values. Only if a given coherence strength does not tie country means within a limited range, could one judge the same country means as in-equivalent, at least as concerns their underlying coherence. But insofar as a given coherence strength ties country means within a limited range, in-equivalence between means can at most be partial, not complete. And that partiality might cover only a minor section of the total variation, in which case it is less of a problem.

The second, and more important, question is whether the existing coherence variability actually matters. It only would if its existence is a source of disturbance that obscures the effects that we theoretically expect from country means in emancipative values. Only if this is the case, could one infer incomparability from variability in coherence.

The third question is to what extent coherence variability is erratic or systematic. Only if the coherence variability is erratic, can it be classified as noise that undermines comparability. If, however, this variability maps in systematic fashion on other aspects of reality, differences in coherence strength unfold on a common basis--the essence of comparability.

Figure 1. Country Means and Coherence Strength in Emancipative Values



Notes: Data are from WVS, waves 4 to 6, taking the latest available survey from each country. Temporal coverage varies from 2000 to 2012, depending on when the latest survey has been conducted. The Emancipative Values Index (EVI) is measured as described in Welzel (2013: 69-73). Vertical axis shows per country the arithmetic population mean in these values; horizontal axis shows per country the Cronbach’s alpha with respect to the four sub-components of the EVI.

Let us address these issues point by point. Figure 1 plots country means in emancipative values against the differential coherence of these values per country (as indicated by the Cronbach’s alpha).²² It is obvious that country means are higher when the coherence of emancipative values is stronger, which is logical: emancipative values can become a prevalent orientation only to the extent to which support for one domain of

²² For each country sample we calculate the Cronbach’s alpha across the four sub-indices of emancipative values: the choice, voice, equality and autonomy index. Aléman/Woods criticize the use of alpha scores because these scores increase with the number of items over which they are calculated. However, this source of variation is controlled here as we use the same set of items in each country. Over a constant set of items, alpha scores are definitely comparable: higher scores indicate a stronger average correlation among the constituent items--a reliable measure of inter-item coherence. We experimented with alternative measures of inter-item coherence, comparing various goodness of fit indices taken from country-wise confirmatory factor analyses. Yet, the pattern of cross-national variation in inter-item coherence is the same in these instances as it is with alpha scores.

these values converges with support for the others. Hence, the same country mean in emancipative values can represent different coherence strengths, yet these differences are tied to a clearly limited range. In other words, in-equivalence with respect to the same country means' coherences is a minor phenomenon. In numbers, in-equivalence accounts for just 37 percent of the variation. This proportion results directly from the fact that 63 percent of the variation in means and coherence coincide: for the major part, higher country means also imply stronger coherence. In a nutshell, differences in coherence strength are mostly incorporated in mean differences.

The next consideration is whether the country means in emancipative values continue to associate with their expected correlates, even if we take into account the limited variation in coherence strength at each mean level. To test this possibility, we regress an expected correlate of emancipative values on the country means in these values, under control of the cross-national variability in these means' coherence.

The expected correlate of our choice here is the "effective democracy index" (EDI). The EDI is a refined measure of democracy that downgrades Freedom House's "civil liberties" and "political rights" ratings for deficiencies in law enforcement that these ratings do not cover but which are tapped by the World Bank's "rule of law" and "control of corruption" scores.²³ Our theory posits that country means in emancipative values predict the countries' EDI scores fairly well. Indeed, Model 1 in Table 1 shows that country means in emancipative values explain 68 percent of the cross-national variance in effective democracy. Now, the question is whether this effect exists independent of variability in these values' coherence and, accordingly, persists under control of this variability. As Model 3 in Table 1 shows, this is beyond doubt indeed the case. In fact, the inclusion of the coherence variability does not really add much to the explained variance in effective democracy (also not as an interaction term: see Model 4) – which is not so surprising when one knows that country means anyways capture most of the variability in coherence strength. Similar results are obtained for many other expected correlates of emancipative values.²⁴ Consequently, there is no in-equivalence in country means as concerns their association with other aspects of reality.

²³ The idea is to measure not just democracy but democracy on the condition that its legal features are put into real practice by law enforcement. For conceptual issues and measurement details see Welzel and Inglehart (2006), Alexander and Welzel (2011a) as well as Alexander, Inglehart and Welzel (2012). The latter publication in particular defends the concept of effective democracy against previous criticism, demonstrating that no other measure of democracy associates stronger than the effective democracy index with democracy's theoretically expected correlates, from economic modernization to social equality to internal and external peace.

²⁴ Detailed results are available upon request from the authors.

Table 1. Regressing Effective Democracy on Country Means in Emancipative Values and their Coherence

<i>Dependent Variable: Effective Democracy Index (EDI) 2012</i>				
<i>Predictors:</i>	Model 1	Model 2	Model 3	Model 4
Constant	0.41***	0.42***	0.42***	0.40***
EVI: Mean ^{a)}	2.25 (0.83)***		1.65 (0.61)***	1.57 (0.58)***
EVI: Coherence ^{b)}		1.31 (0.77)***	0.52 (0.36)***	0.54 (0.37)***
Mean X Coherence ^{c)}				1.27 (0.13)
Adj. R ²	0.68	0.59	0.74	0.74
N (countries)	100	97	97	97

Notes: Entries are unstandardized regression coefficients with partial correlations in parentheses. Test statistics for heteroskedasticity (White test), collinearity (variance inflation factors) and influential cases (DFFITs) indicate no violation of OLS assumptions. The dependent variable is Alexander, Inglehart and Welzel's (2012) Effective Democracy Index, updated for 2012 and rescaled so that the theoretical minimum is 0 and the maximum 1, with fractions for intermediate positions. Independent variables are taken from waves 4 to 6 of the WVS, using the latest survey for each country. Thus, temporal coverage varies from 2000 to 2012. Significance levels: *** P < 0.001; ** P < 0.010; * P < 0.050.

^{a)} EVI stands for Emancipative Values Index, as defined by Welzel (2013: 69-74). "EVI: Mean" measures per country the arithmetic population mean in these values.

^{b)} "EVI: Coherence" measures per country the Cronbach's alpha with respect to the four constituent sub-indices of emancipative values.

^{c)} "Mean X Coherence" is a multiplicative interaction term between "EVI: Mean" and "EVI: Coherence." To build this term, "EVI: Mean" and "EVI: Coherence" have been centered on their arithmetic means and have been introduced in this rescaled format in Model 4.

The third consideration points to the forces that shape the coherence in emancipative values. Considering this topic is at the same time a way to address the issue of cultural bias in measurement – a point of outstanding importance in cross-national survey research.

THE ISSUE OF CULTURE BIAS

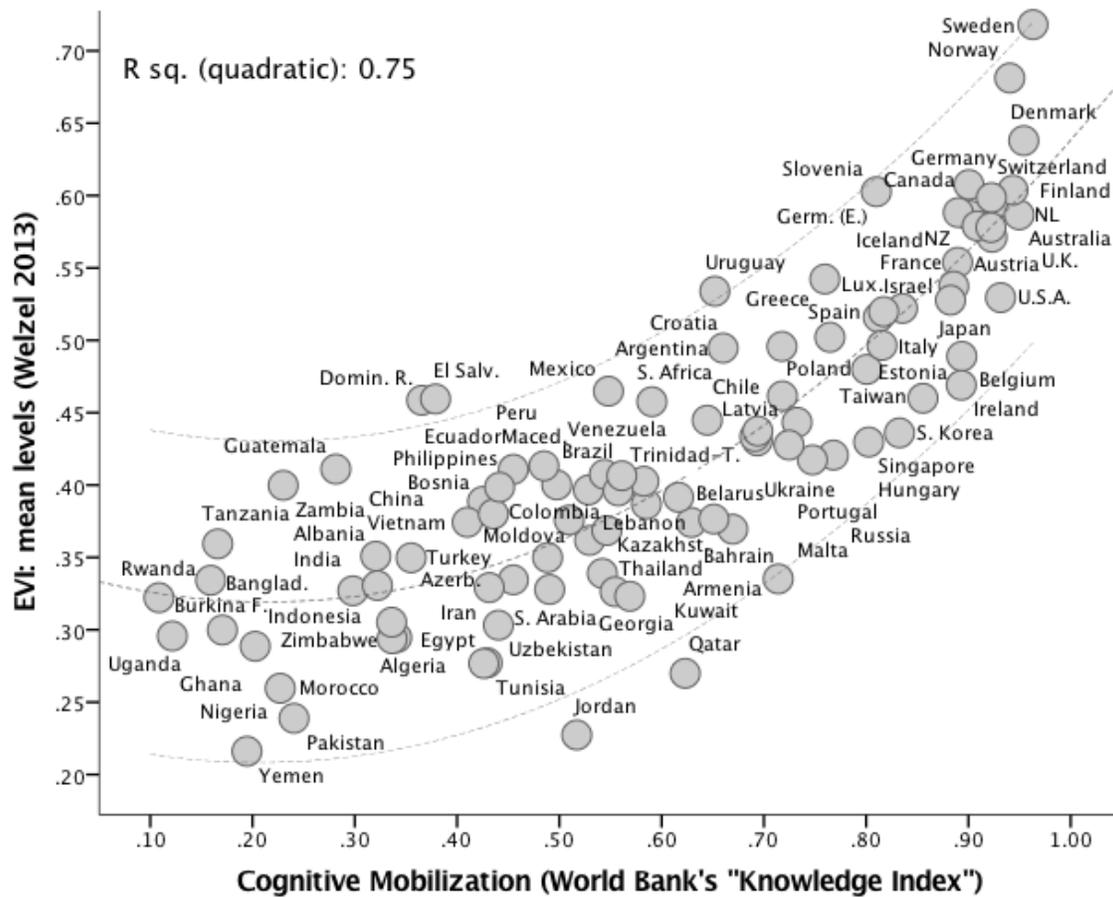
Culture zones differ significantly in their mean levels of emancipative values (Welzel 2013: 89). The most opposite positions are taken by historically Protestant countries in Western Europe, which score exceptionally high in emancipative values, and Islamic countries in the Middle East and South Asia, which score outstandingly low in these values (Alexander/Welzel 2011b). Yet, the simple fact that country scores in emancipative values map closely on culture zones cannot as such be interpreted as a disqualifying cultural bias in measurement. On the contrary, the mapping pattern underlines the construct's quality in highlighting cultural differences. And these differences are meaningful because the countries of those culture zones that perform particularly weak in emancipative values also lack some of the most fundamental correlates of these values, such as cognitive mobilization and effective democracy.

Another take on this problem is to examine the social forces that shape cross-country differences in inter-item coherence. If we can identify such forces and show that their presence at the country level systematically strengthens the individual-level coherence in emancipative values, then variability in this coherence can, again, not be taken as an indication of incomparability. For the variability maps on a common reference standard--the very basis of comparability.

Yet another question is to what extent cultural traditions belong to these coherence-shaping forces. Insofar as they do, culture zone differences need to be seen as impermeable and persistent. If, however, social forces related to modernization induce coherence into emancipative values, cultural differences can diminish under the imprint of modernization.

A formidable candidate for a coherence-inducing force is cognitive mobilization. A pivotal aspect of modernization, cognitive mobilization advances through expanding education, skill specialization, widening access to information, technological progress and greater intellectual stimuli in people's daily activities. All of these are aspects of rising knowledge societies. The rise of this new type of society has increased people's cognitive capacities, as is evident in rising IQ-levels—the so called "Flynn effect" (Trahan et al. 2014). Flynn (2012) himself interprets this effect as indicating a general rise in cognitive capacities, triggered by the cognitive impulses of emerging knowledge societies. Resonating with Pinker's (2011) "escalator of reason," Flynn speculates that cognitive mobilization also enhances people's moral judgment capacities, improving their understanding of universal ethical principles, such as those related to human emancipation.

Figure 2. Cognitive Mobilization and Emancipative Values



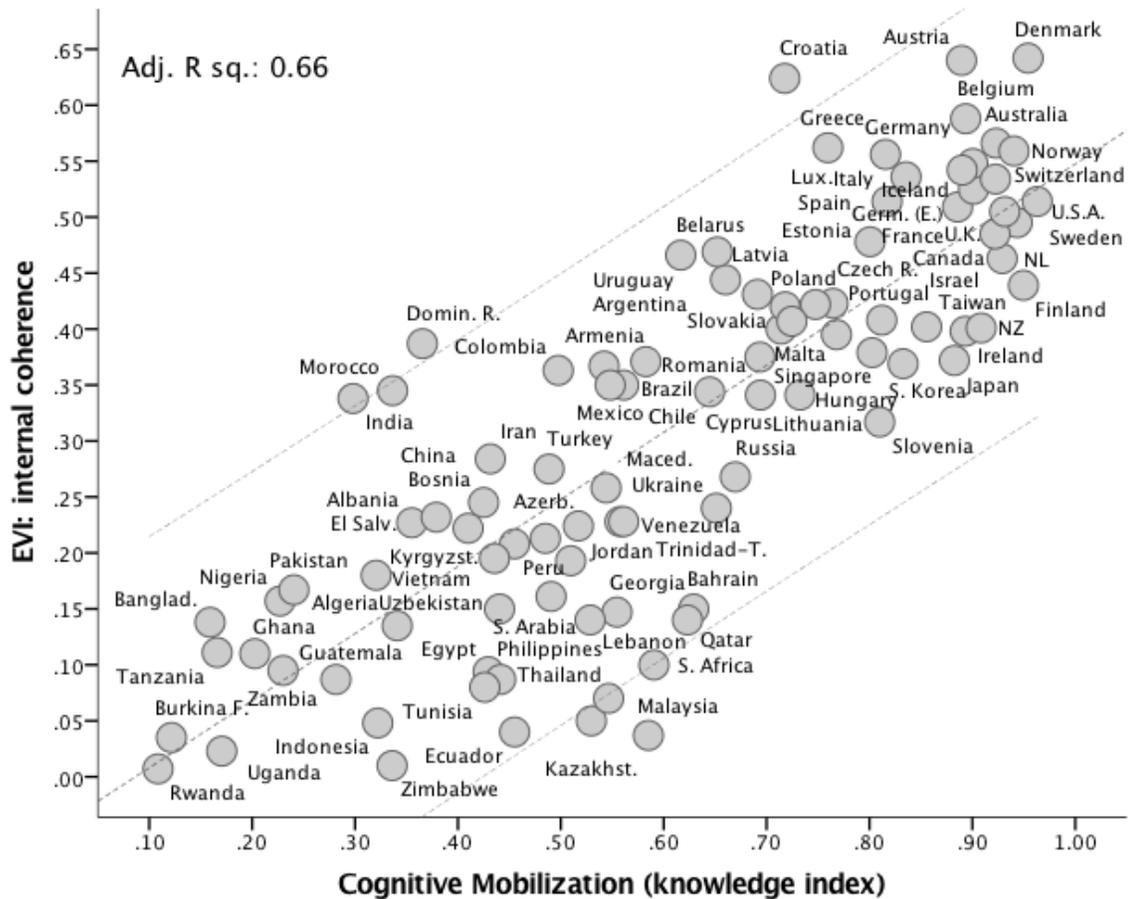
Notes: Horizontal axis measures the countries' advancement in cognitive mobilization, using a rescaled version of the World Bank's "Knowledge Index (KI)," as detailed in Welzel (2013: 18). The index summarizes country-level information on the spread of information technology, educational achievement and per capita scientific output. Temporal coverage varies from 1995 to 2005. Vertical axis measures per country the arithmetic population mean in the Emancipative Values Index (EVI) as detailed in Welzel (2013: 69-73). Data are from WVS, waves 4 to 6, taking the latest available survey from each country. Temporal coverage varies from 2000 to 2012, depending on when the latest survey has been conducted.

Confirming this assumption, cognitive mobilization explains fully 75 percent of the cross-national differences in emancipative values, as Figure 2 illustrates: societies that are more advanced in cognitive mobilization also tend to have more prevalent emancipative values.

What is more, cognitive mobilization also enhances the coherence in people's moral judgment. This is obvious from Figure 3, which plots the individual-level coherence in emancipative values per country against each country's degree of cognitive mo-

bilization. We see a strongly linear distribution, suggesting that emancipative values become more coherent as a country's cognitive mobilization advances.²⁵

Figure 3. Cognitive Mobilization as a Coherence-inducing Force in Emancipative Values



Notes: Horizontal axis measures the countries' advancement in cognitive mobilization, using a rescaled version of the World Bank's "Knowledge Index (KI)," as detailed in Welzel (2013: 18). The index summarizes country-level information on the spread of information technology, educational achievement and per capita scientific output. Temporal coverage varies from 1995 to 2005. Vertical axis shows per country the Cronbach's alpha with respect to the four sub-components of the Emancipative Values Index (EVI). Data are from WVS, waves 4 to 6, taking the latest available survey from each country. Temporal coverage varies from 2000 to 2012, depending on when the latest survey has been conducted.

²⁵ One can show for various item batteries in the WVS, such as that on confidence in institutions, that individual-level correlations within countries are stronger when the respective populations are cognitively more mobilized.

Table 2. Regressing the Coherence of Emancipative Values on Cognitive Mobilization, Democratic Traditions and Global Linkages

<i>Dependent Variable:</i> Individual-level Coherence of Emancipative Values per Country							
<i>Predictors:</i>	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Constant	-0.04	0.15***	0.07	-0.04	-0.04	-0.06	-0.04
Cognitive Mobilization 1995	0.58 (0.80)***			0.55 (0.71)** *	0.52 (0.51)** *		0.51 (0.50)* **
Democratic Traditions 1900-1995		0.33 (0.54)** *		0.04 (0.09)		0.08 (0.14)	0.04 (0.07)
Global Linkages 1995			0.69 (0.69)** *		0.08 (0.07)	0.59 (0.51)** *	0.05 (0.04)
Adj. R ²	0.64	0.29	0.47	0.63	0.60	0.47	0.60
N (countries)	96	98	85	96	84	85	84

Notes: Entries are unstandardized regression coefficients with partial correlations in parentheses. Test statistics for heteroskedasticity (White test), collinearity (variance inflation factors) and influential cases (DFFITs) indicate no violation of OLS assumptions. The dependent variable measures per country the Cronbach's alpha with respect to the four constituent sub-indices of emancipative values, as defined by Welzel (2013: 69-73). Cognitive mobilization is measured using a rescaled version of the World Bank's "Knowledge Index (KI)." Democratic Traditions are measured using a rescaled version of Gerring et al.'s (2005) "democracy stock" variable. Global linkages are measured using Dreher et al.'s (2008) scores of a country's integration into global economic, social, cultural, and political exchange. All predictors have a scale range from minimum 0 to maximum 1, with fractions of 1 indicating intermediate positions. Significance levels: *** P < 0.001; ** P < 0.010; * P < 0.050.

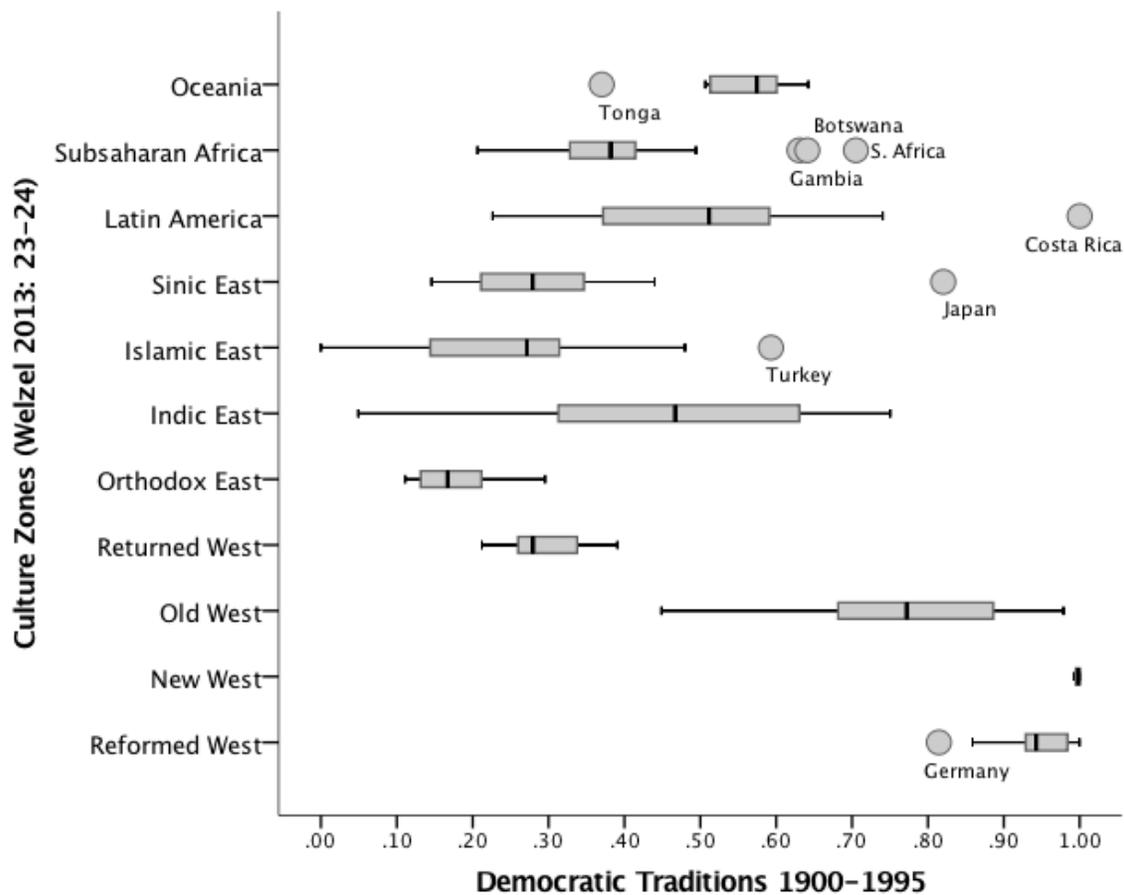
Interestingly, cognitive mobilization eliminates the impact that cultural traditions seem to exert before we take cognitive mobilization into account. This is evident from Table 2 where we regress the coherence of emancipative values per country on the countries' cognitive mobilization and their democratic tradition, using Gerring et al.'s (2005) "democracy stock" variable.²⁶

The democratic tradition is a first-rate measure of cultural traditions more generally speaking: 72 percent in the variance of this measure across 188 countries is due to

²⁶ "Democracy stock" measures a country's accumulated historic experience with democracy from 1900 to 1995, as detailed in Gerring et al. (2005).

differences between the ten culture zones defined by Welzel (2013: 89). The boxplot in Figure 4 demonstrates this point and the evidence is not surprising: because democracy is the signature feature of Western culture, its very endurance indicates how early and deeply cultures around the world have been “infiltrated” with Western values. From the viewpoint of institutional learning, it is highly plausible that persistent democratic socialization over many generations induces coherence into emancipative values—since these values' emphasis on freedom of choice and equality of opportunities is at the heart of the whole idea of democracy.

Figure 4. Global Variation in Democratic Traditions across Culture Zones



Notes: Democratic Traditions are measured using a rescaled version of Gerring et al.’s (2005) “democracy stock” variable. Data cover 188 countries. Countries are assigned to culture zones as defined by Welzel (2013: 23-24). The eleven culture zones account for 72 percent of the entire cross-national variation in democratic traditions.

It is equally plausible that a country’s exposure to “global justice scripts” induces coherence into emancipative values because these scripts often invoke emancipatory

ideals. As a proxy for such exposure, we use Dreher et al.'s (2008) measures of global linkages in Table 2.

Now, the regressions in Table 2 resolve these issues quite clearly: (1) cultural bias, manifest in the strength of the democratic tradition, shows no more effect on the coherence of emancipative values in a country, once we take cognitive mobilization into account; (2) the same is true for global linkages: no effect after controlling cognitive mobilization; (3) the latter, by contrast, powerfully induces coherence into emancipative values—fully irrespective of the democratic tradition and global linkages. Accordingly, coherence is induced into emancipative values by a key aspect of modernization. This finding underlines the validity of emancipative values as a measure of modernization's manifestation in collective mentalities.

Nevertheless, there are reasons to remain skeptical. Perhaps, many people's responses to survey questions on matters of emancipation are meaningless because these people do not understand what emancipation is about.²⁷ All the more, this might be the case when the respective issues are not controversial in a society, reflecting a tacit consensus on traditional morality.

As plausible as this suspicion might appear at first glance, it is untenable upon closer examination. The WVS does not ask respondents for their position on emancipation in an abstract sense, using some form of incomprehensible academic language. Instead, the WVS addresses such down-to-earth topics as male dominance, child obedience and heterosexual norms, using rather colloquial terminology. It is hard to believe that people have no first-hand experience with such everyday themes: these issues are integral parts of what might be described as "evolutionary normality" in moral systems.

For centuries and millennia, male dominance, child obedience and strict heterosexuality have been the norm throughout most human societies. Taking the opposite—emancipatory—positions on these issues is an "evolutionary novelty" that signals a breakup of traditional limitations on human morality (Alexander/Inglehart/Welzel 2015). Where societal conditions have not matured to this point, respondents will naturally stick to the evolutionary norm and take traditional moral positions on questions of emancipation. If so, we will inevitably obtain a low score in emancipative values, which tells us something very real: how little appeal emancipatory ideals have in a given society.

Whether or not the majority of a given society wishes to be measured against the standards of emancipation is a different question. But this question should not concern researchers when a society's performance on these standards has predictable consequences for such important things as human rights, democracy, peace and sustainability.

Especially in morally conservative societies, it might well be that some respondents indicate anti-emancipatory positions because the social norm pressures them to do so. If

²⁷ There certainly is much variability in how respondents understand items related to emancipative values. Thus, we face semantic inequivalence among individual-level responses (cf. van Deth 2013). But apparently, much of the semantic inequivalence is a randomly distributed phenomenon and, hence, cancelled out through aggregation. Otherwise, country-level scores in emancipative values could impossibly map so tightly on other country-level characteristics, especially those of an objective nature.

so, the individual-level score in emancipative values would not measure a true preference for a certain percentage of respondents. Yet, the country-level score would still be a valid measure because it indicates accurately the strength of anti-emancipatory norms--an important aspect of reality, given that social norms have considerable power in shaping human behavior.

CONCLUSION

Advocates of structural equation modelling judge multi-item constructs against three standards: (1) multiple items converge in a single dimension; (2) individual-level patterns of item convergence are invariant across aggregate-level units; (3) aggregate-level patterns of item convergence replicate those at the individual level.

This set of requirements involves two far-reaching premises: measurement validity hinges solely on a construct's internal convergence and convergence patterns at the individual level have priority over those at the aggregate level.

We have argued that both premises are profoundly flawed for a couple of reasons. To begin with, convergence patterns in the aggregate exist in their very own right, have real consequences and are more clearly structured than convergence patterns at the individual level--the level where measurement error is abundant. Also, societies are powerfully shaped by ecological effects for which no individual-level equivalents exist. Consequently, the premise of ontological primacy of the individual level over the aggregate level is mistaken.

Next, internal convergence is certainly a point of consideration to describe the structure of multi-item constructs. But external linkage is another aspect and one that tells us more about a construct's outreach into reality, including its predictedness, predictiveness and explanatory power. As concerns combinatory constructs, external linkage definitely trumps internal convergence as a validity criterion because these constructs explicitly allow for multi-dimensionality, in which case internal convergence is not an even requirement. Thus, the priority of internal convergence is a fallacious premise too.

Since both premises are flawed, the SEM-approach has no explanation for an overlooked but frequently occurring phenomenon: a construct's lacking convergence at the individual level goes hand in hand with powerful and important linkages at the aggregate level.

Still, there is a perfectly logical explanation of this phenomenon--component substitutability: the same aggregate-level scores can emerge from different mixtures and overlaps among component scores at the individual level, and yet these aggregate-level scores map in similar fashion on a construct's expected antecedents and consequences. The reason: variable compositions at the individual level are mutually substitutable when they produce the same aggregate-level score. Since component substitutability is the anti-thesis of internal convergence, it is beyond the comprehension of the SEM-approach.

Aléman/Woods' criticism of emancipative values and other constructs from the WVS is based on the premises of the SEM-approach and mistaken for this reason. If we

were seeking to identify orientational coherence patterns inside individuals, Aléman/Woods' critique would have a point. But that very explicitly is not our goal. Instead, our constructs intend to capture patterns in value orientations that emerge first and foremost, and at times solely, in the aggregate. Only value patterns that exist in the aggregate can have an impact on other systemic phenomena of some relevance, such as human rights, democracy, peace, and sustainability. As long as we are interested in such outcomes, we should continue to study aggregate patterns in values. At the same time, we should stop judging these aggregate patterns by whether or not they are replicated at the individual level. The aggregate is a reality level in its own right.

In conclusion, we advocate a decidedly "nomological" view: constructs should be judged valid whenever the same overall scores map in corresponding fashion on expected antecedents or consequences. Scholars should consider something as real when it shows up as real in its preconditions and outcomes.

REFERENCES

- Aléman, J. & D. Woods (2015). "Value Orientations from the World Values Survey: How Comparable are They Cross-nationally?" *Comparative Political Studies* 48: 1-29 (DOI: 10.1177/0010414015600458).
- Alexander, A. & C. Welzel (2011a). "Measuring Effective Democracy: The Human Empowerment Approach." *Comparative Politics* 43: 271-289.
- Alexander, A. & C. Welzel (2011b). "Islam and Patriarchy: How Robust is Muslim Support for Patriarchal Values?" *International Review of Sociology* 21: 249-275.
- Alexander, A., R. Inglehart & C. Welzel (2012). "Measuring Effective Democracy: A Defense." *International Political Science Review* 33: 41-62.
- Alexander, A.C., R. Inglehart & C. Welzel (2015). "Sexual Emancipation: Breakthroughs into a Bulwark of Tradition." *Social Indicators Research* 102: forthcoming.
- Coltman, T., T.M. Divenney, D.F. Midgley & S. Venaik (2008). "Formative versus Reflective Measurement Models." *Journal of Business Research* 61: 1250-1262.
- Dahlum, S. & C.H. Knutsen (2015). "Democracy by Demand?" *British Journal of Political Science* 46: forthcoming.
- Davidov, E., H. Duellmer, E. Shlueter, P. Schmidt & B. Meulemann (2012). "Using a Multilevel Structural Equation Modeling Approach to Explain Cross-National Measurement Nonvariance." *Journal of Cross-cultural Psychology* 43: 558-575.
- Delhey, J., K. Newton & C. Welzel (2011). "How General is Trust in 'Most People'? Solving the Radius-of-Trust Problem." *American Sociological Review* 76: 786-807.
- Dreher, A., N. Gaston & W.J.M. Martens (2008). *Measuring Globalization*. New York: Springer.
- Flynn, J.R. (2012). *Are We Getting Smarter? Rising IQs in the Twenty-first Century*. New York: Cambridge University Press.
- Gerring, J., P. Bond, W.T. Barndt & C. Moreno (2005). "Democracy and Economic Growth." *World Politics* 57: 323-64.

- Hofstede, G. (2001 [1980]). *Culture's Consequences*. Beverly Hills: Sage.
- Inglehart, R. (1990). *Culture Shift in Advanced Industrial Societies*. Princeton: Princeton University Press.
- Inglehart, R. (1997). *Modernization and Postmodernization*. Princeton: Princeton University Press.
- Inglehart, R. & P. Norris (2003). *Rising Tide*. New York: Cambridge University Press.
- Inglehart, R. & C. Welzel (2003). "Political Culture and Democracy: Analyzing Cross-level Linkages." *Comparative Politics* 36: 61-79.
- Inglehart, R. & C. Welzel (2005). *Modernization, Cultural Change and Democracy: The Human Development Sequence*. New York: Cambridge University Press.
- Inglehart, R. & C. Welzel (2010). "Changing Mass Priorities: The Link between Modernization and Democracy." *Perspectives on Politics* 8: 551-567.
- Inglehart, R., P. Norris & C. Welzel (2002). "Gender Equality and Democracy." *Comparative Sociology* 1: 235-264.
- Kaase, M. (1986). „Das Mikro-Makro-Puzzle der empirischen Sozialforschung: Anmerkungen zum Problem der Aggregatstabilität bei individueller Instabilität in Panelbefragungen." *Koelner Zeitschrift fuer Soziologie und Sozialpsychologie* 38: 209-222.
- Leung, K., & M.H. Bond (1989). "On the Empirical Identification of Dimensions for Cross-cultural Comparisons." *Journal of Cross-Cultural Psychology* 20: 133-152.
- Leung, K., & M.H. Bond (2007). "Psycho-logic and Eco-logic: Insights from Social Axiom Dimensions." In van de Vijver, F., D. van Hemert & Y. P. Poortinga (eds.), *Individuals and Cultures in Multilevel Analysis*. Mahwah: Lawrence Erlbaum Associates, pp. 199-221.
- Norris, P. & R. Inglehart (2004). *Sacred and Secular*. New York: Cambridge University Press.
- Ones, D. & C. Viswesvaran (1996). "Bandwidth-Fidelity-Dilemma in Personality Measurement for Personnel Selection." *Journal of Organizational Behavior* 17: 609-26.
- Page, B.I. & R.Y. Shapiro (1992). *The Rational Public*. Chicago: University of Chicago Press.
- Pinker, S. (2011). *The Better Angels of Our Nature*. London: Allen Lane.
- Przeworski, A. & H. Teune (1970). *The Logic of Comparative Social Inquiry*. New York: Wiley.
- Spaiser, V., S. Ranganathan, R.P. Mann & D.J.D. Sumpter (2014). "The Dynamics of Democracy, Development and Cultural Values." *PLoS ONE* 9: 1-11.
- Spaiser, V. & D.J.D. Sumpter (2015). "Testing and Revising the Human Development Sequence Theory using Agent-Based Simulation and Data." Online Manuscript available at ResearchGate (DOI: 10.13140/RG2.1.4535.0244).
- Stegmüller, D. (2011). "Apples and Oranges? The Problem of Equivalence in Comparative Research." *Political Analysis* 19: 471-487.
- Stimson, J.A., M.B. MacKuen & R.S. Erikson (2002). *The Macro Polity*. New York: Cambridge University Press.

- Trahan, L., K.K. Stuebing, M.K. Hiscock & J.M. Fletcher (2014). "The Flynn-Effect: A Meta-analysis." *Psychological Bulletin* 140: 1332-1360.
- van de Vijver, F.J.R. (2011). "Capturing Bias in Structural Equation Modelling." In Davidov, E., P. Schmidt & J. Billiet (eds.), *Cross-Cultural Analysis*, London: Routledge, pp. 3-34.
- van Deth, J. (2013). "Equivalence in Comparative Political Research." In J. van Deth (ed.), *Comparative Politics: The Problem of Equivalence*. London: ECPR Press, pp. 1-19.
- Vanhanen, T. (2003). *Democratization*. London: Routledge.
- Welzel, C. (2007). "Are Levels of Democracy Influenced by Mass Attitudes?" *International Political Science Review* 28: 397-424.
- Welzel, C. (2006). "Democratization as an Emancipative Process." *European Journal of Political Research* 45: 871-896.
- Welzel, C. (2013). *Freedom Rising: Human Empowerment and the Quest for Emancipation*. New York: Cambridge University Press.
- Welzel, C. & R.J. Dalton (2014). "From Allegiant to Assertive Citizenship." In Dalton, R.J. & C. Welzel (eds.), *The Civic Culture Transformed: From Allegiant to Assertive Citizens*, New York: Cambridge University Press, pp. 282-306.
- Welzel, C. & F. Deutsch (2011). "Emancipative Values and Nonviolent Protest: The Importance of 'Ecological' Effects." *British Journal of Political Science* 42: 465-479.
- Welzel, C. & R. Inglehart (2006). "Emancipative Values and Democracy." *Studies in Comparative International Development* 41: 74-94.
- Welzel, C. & R. Inglehart (2010). "Values, Agency, and Well-Being: A Human Development Model." *Social Indicators Research* 97: 43-63.
- Welzel, C., R. Inglehart & H.-D. Klingemann (2003). "The Theory of Human Development: A Cross-Cultural Analysis." *European Journal of Political Research* 42: 341-380.
- Welzel, C., R. Inglehart & F. Deutsch (2005). "Social Capital, Voluntary Associations, and Collective Action: Which Aspects of Social Capital Have the Greatest 'Civic' Payoff?" *Journal of Civil Society* 1: 121-146.
- Welzel, C., R. Inglehart & S. Kruse (2015). "Pitfalls in the Study of Democratization." *British Journal of Political Science* 46: forthcoming.