NEW MEASUREMENT INSTRUMENT

Open Access

Measuring public knowledge on nuclear weapons in the post-Cold War: dimensionality and measurement invariance across eight European countries

Fabrício M. Fialho[*]

Abstract

Research on public opinion and international security has extensively examined attitudes toward nuclear weapons, but the diffusion of basic knowledge about nuclear weapons among the everyday citizens has nevertheless been mostly missed. This study proposes a working definition and advances a measurement model of knowledge on nuclear weapons in the general public. It analyzes data from two novel surveys conducted in 2018 ($N = 6559$) and 2019 ($N = 6227$) where respondents from Belgium, France, Germany, Italy, the Netherlands, Poland, Sweden, and the United Kingdom answered a web survey on attitudes and factual knowledge on nuclear weapons. Exploratory and confirmatory factor analytic models are used to examine the dimensionality and to assess the measurement invariance of a scale of knowledge about nuclear weapons. A bifactor measurement model, where a strong general factor represents the construct of interest and specific factors account for the presence of testlets due to questionnaire design, is established and validated. Configural, metric, and scalar invariance are established across the eight samples. The findings indicate that knowledge about nuclear weapons in the general, non-expert public can be reliably measured cross-nationally.

Introduction

Scholarship on public opinion and international security have paid close attention to attitudes toward nuclear weapons and concerns about nuclear war, their impact on mental health, their importance for security, and support to policy decisions during the Cold War and afterwards (e.g., Boehnke et al., 1989; Egeland & Pelopidas, 2021; Fiske et al., 1983; Haste, 1989; Haworth, Sagan, & Valentino 2019; Herron & Jenkins-Smith, 2006, 2014; Herzog & Baron, 2017; Kramer et al., 1983; McAllister & Mughan, 1986; Pelopidas, 2017; Press et al., 2013; Rosi, 1965; Russett, 1990-1991; Russett & Deluca, 1983; Sagan & Valentino, 2017; Zweigenhaft, 1984; Zweigenhaft et al., 1986). The literature on public's perceptions about nuclear weapons has nevertheless almost entirely missed what the public knows about such weapons.

Compared with what has been learned about attitudes toward nuclear weapons, public knowledge about such devices remains lesser studied and known. Since Graham's (1988) review of measures of public knowledge of nuclear issues from then-existing surveys, little progress has been made. For instance, Pierce et al. (2000) examined perceived familiarity with terms related to nuclear weapons production rather than factual knowledge in their comparative study in areas in Russia and the USA. In their study of Indian elites, Cortright and Mattoo (1996) found that respondents believed to be difficult to obtain information about nuclear weapons in

*Correspondence: f.mendes-fialho@lse.ac.uk
International Inequalities Institute, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

the country, but no measure of factual knowledge is presented. The paucity of studies on public knowledge about nuclear weapons leaves important gaps in the fields of public opinion and international security.

Knowledge is a critical political asset. It has been shown to be a strong indicator of political awareness and that informed citizens are better equipped to identify relevant political events and actors, to understand rules and regulations, and to evaluate political choices (Delli Carpini & Keeter, 1996; Zaller, 1992). A better informed citizenry is more responsive, responsible, and better posited to keep political elites accountable. Despite the importance of nuclear weapons in international politics for the human, strategic, and financial costs involved, scholars have not attempted to document how well-equipped citizens are to understand events and to make political judgments and choices on the matter. Public understanding of nuclear weapons politics becomes even more relevant as nuclear-armed states have been investing in programs to prolong the operational life of their nuclear weaponry (Kristensen, 2014). A stronger understanding of what citizens factually know about the nuclear weapons world would be useful for many purposes, from the documentation of what aspects are broadly known and what only those passionate about the topic comprehend, to inform educational campaigns, and to understand how preferences and values on nuclear weapons affairs vary across levels of information about those weapons, to name just a few. A solid understanding of the public's factual knowledge is therefore key to also assess their attitudes and preferences on the matter.

I analyze data from two novel public opinion surveys carried in eight European countries to contribute to the literature on the measurement of public's knowledge on nuclear weapons with the assessment of dimensionality and measurement invariance of a scale that focuses on "structural" aspects of nuclear weapons rather than on awareness of salient topics of the moment. Results demonstrate the feasibility of measuring knowledge about nuclear weapons among the public and its cross-national comparability. Overall, I find that that measurement of "general" and "static" aspects of nuclear weapons politics as defined in the next section comprise a reliable scale with sound psychometric properties even if measured with relatively few items and is capable of detecting cross-national differences in latent means and variance of the construct.

In what follows, a working definition of public knowledge about nuclear weapons, or "nuclear knowledge", is provided; an argument for the (essentially) unidimensionality of the construct is presented. The data and methods used in the paper are then presented. Item

analysis and item selection are followed by an assessment of the dimensionality using the calibration sample; the establishment of a bifactor model that preserves the "essential unidimensionality" of the construct of interest is discussed, and a measurement invariance test across eight countries is performed. The measurement model is then replicated using a validation sample. The criterion validity of the scale is assessed and recommendations on its use are provided. The paper concludes with suggestions for future research.

Public knowledge about nuclear weapons: towards a working definition

Public opinion research and the nuclear weapons scholarship have paid scanty attention to what individuals know about nuclear weapons. Previous studies have focused either on *perceptions of familiarity* with nuclear weapons-related terms and of availability of relevant information without directly measuring what respondents know about them (e.g., Cortright & Mattoo, 1996; Pierce et al., 2000), or on documenting *awareness* of— i.e., whether a respondent has or had "heard about"— the most recent political developments in the field such as the signing of an international treaty on nuclear weapons (Graham, 1988). Although scholars have been encouraged to "move beyond the idea that the public is poorly informed", to "study patterns of knowledge and awareness", and to "concentrate on identifying the subtle relationship between knowledge and attitudes" (Graham, 1988, p. 321), little progress has unfortunately been made conceptually and empirically. The absence of a definition of what information or *knowledge* about nuclear weapons means and how to measure it has delayed progress in the field vis-a-vis the rich literature on *attitudes* toward nuclear weapons (see, among others, Fiske et al., 1983; Haworth et al., 2019; Herron & Jenkins-Smith, 2006; Kramer et al., 1983; Press et al., 2013; Sagan & Valentino, 2017; Zweigenhaft et al., 1986). A working definition of knowledge about nuclear weapons, even if provisional, must be laid down.

Given the unique status of nuclear weapons in international *politics*, a working definition of knowledge about nuclear weapons may benefit from the well-established literature on *political* knowledge. Delli Carpini and Keeter (1996), p. 10, 294) define political knowledge as "the range of factual information about politics that is stored in long-term memory" (see also Barabas et al., 2014). Barabas et al. (2014) propose a typology where knowledge on political objects can be organized along two dimensions, a temporal dimension and a topical dimension. The temporal dimension

accounts for how recently a fact commences or is established, and can be schematically divided in "surveillance" (recent developments that might be learned from monitoring mass media)[1] and "static" (facts established and in circulation for a long time, eventually incorporated in the education system, documentaries, publications, and so on). The topical dimensions pertains to the type of fact, whether it has to do with policy issues (specific scope) or with political institutions and players (general scope). Altogether, political knowledge refers to the retention of factual information on recent or older events and developments related to policies or political institutions and players. This definition, as I argue next, can be transferred to the realm of nuclear weapons.

Knopf (2012, p. 81) claims that there are facts and information about nuclear weapons that are "well established and more or less objective and incontrovertible" thus "acquiring knowledge about these facts is therefore factual learning"—an argument that closely resembles the very definition of political knowledge and serves as a solid building block for a definition of public knowledge about nuclear weapons.

Considering the particularities of nuclear weapons politics, it is of interest to consider how the temporal and topical dimensions of political knowledge contribute to the definition and operationalization of public knowledge about nuclear weapons. Although individuals' attitudes toward foreign and defense policy, nuclear weapons included, have a stable structure and awareness about them are relatively widespread among the public (Eichenberg, 1998; Graham, 1988; Herron & Jenkins-Smith, 2014; Knopf, 2012), the relative salience of nuclear weapons issues is overall low and rarely rank among the top policy priorities of survey respondents (Cortright & Mattoo, 1996; Flynn & Rattinger, 1985; Schuman et al., 1986; Wilson, 2015). Public concerns and activism on the matter dramatically fluctuates in time; transient moments of heightened interest tend to follow international crisis and vanish afterwards (Kramer et al., 1983; Schuman et al., 1986; Wilson, 2015). With the exception of specialized issue publics (Iyengar, 1990; Krosnick, 1990) and nuclear weapons aficionados, it would be unrealistic to expect citizens to constantly monitor the media and specialized outlets in search for novel information on nuclear weapons. Knowledge on nuclear weapons politics, except in public opinion studies focused on awareness on the latest international crisis, should rather focus on "static" rather than "surveillance" facts.

One other relevant aspect of nuclear weapons policies is the secrecy that surrounds policy aspects and decision-making processes, which may remain undisclosed for decades, as well as the absence of straightforward policy information available to the public such as costs, deployment, conditions for use of such weapons, and so on. In fact, it has been argued that only a small cadre of high-rank specialists would have full access to policy details and that such information may even be held from political authorities (Dahl, 1985; Ellsberg, 2017; Rosenbaum, 2011). Moreover, it has been also claimed that national security policies are "strongly contested even among policy specialists" (Herron & Jenkins-Smith, 2006, p. 168). The combination of disagreement among specialists and the secrecy on core policy aspects makes nuclear weapons policies opaque and extremely difficult to be tracked by non-specialists. Therefore, unless a said survey is designed specifically to assess public awareness on highly visible policy developments such as the signing of international treaties on nuclear weapons (such as lion's share of the survey items examined by Graham, 1988), survey items on policies seems to be a less-than-optimal choice.

Per the discussion above, I argue that mass public opinion surveys that aim to assess the public's knowledge on nuclear weapons should target *general* and *static* aspects of nuclear weapons politics rather than transient "breaking news" and opaque policy-oriented issues. The domain of knowledge on nuclear weapons would therefore comprise static–general facts (Barabas et al., 2014) and measure "structural" aspects of nuclear weapons politics. It would assess facts that have been established enough time ago for the information to be disseminated and assimilated by individuals in a scenario of sporadic media coverage and low issue salience, with the transmission of information mostly taking place via the education system, TV shows and documentaries, movies and popular culture, and so on. Importantly, the training required for the understanding of science of nuclear devices should rule technical aspects of such weapons out of the measure. For an individual's understanding of politics of nuclear weapons and their implications for international politics, it is argued, for instance, that knowing that Hiroshima and Nagasaki were bombed in World War II using atomic weapons and that the detonation of such devices resulted in massive destruction, dozens of thousands of deaths, and the release of high levels of radiation is more relevant than knowing whether Little Boy and Fat Man employed either fission or fusion technology. Another important definition of the delimitation of the domain of the knowledge on nuclear weapons related to dimensionality concerns. The emphasis on static–general facts on nuclear weapons politics imposes limits on the scope of the construct and theoretically bounds its dimensionality to one. Such a concept is, from a substantial standpoint,

---

[1] Barabas et al. (2014, p. 845) define surveillance facts as those established in the past 100 days.

unidimensional and distinguishable from (yet possibly correlated with) other dimensions of (political) knowledge in Barabas et al. (2014)'s fourfold typology.

Data and methods

Samples

I analyze two independent surveys in the present study. The calibration sample comes from a novel online survey conducted by YouGov in eight European countries in June 2018. The surveyed countries encompass nuclear weapon–possessing states (France, the United Kingdom), countries that host US nuclear weapons (Belgium, Germany, Italy, the Netherlands), a country that started then terminated its native nuclear weapons program (Sweden), and a country set to host anti-ballistic missile batteries in Eastern Europe (Poland). Female and male adults of 18 to 50 years old comprise the target population, and respondents were recruited to match the gender and age composition of each country. Age was capped at fifty for the survey was originally designed to investigate attitudes toward nuclear weapons among individuals who came of age at the later phases of the Cold War and thereafter. Sample sizes are around 1000 respondents in France and the UK and around 750 respondents in the other countries. As a validation sample, I analyze a second cross-national survey on attitudes toward nuclear weapons carried in the same countries in October 2019 by the IFOP polling research firm. Sample sizes in the 2019 survey resemble those in the 2018 study; respondents are female and male adults of 18 years of age or older. In both surveys, respondents answered a questionnaire in the language of their country of residence (or region, in the Belgian case).[2]

Items

The calibration questionnaire contains eight items on knowledge about nuclear weapons that can be organized in six major themes: (i) the atomic bombing in World War 2, its targets and casualties; (ii) nuclear weapons possessors;[3] (iii) effects of a nuclear weapon explosion; (iv) whether any country has ever terminated a nuclear weapons program; (v) number of existing nuclear weapons, in specific countries and in the world; and (vi) number of nuclear weapons tests ever carried. Different item formats are employed. Table 1 presents full question wording, item format, response options (correct

---

[2] YouGov proprietary online opt-in panel was used for recruitment in France, Germany, Italy, Sweden, and the UK; YouGov partners recruited responded and carried the data collection in Belgium, the Netherlands, and Poland. IFOP recruited participants from the Bilendi proprietary online panels.

[3] The nine countries that currently maintain nuclear weapons in their arsenals are: China, France, India, Israel, North Korea, Pakistan, Russia, the UK, and the USA. Five of them—China, France, Russia, the UK, and the USA—are permanent members of the United Nations Security Council.

responses italicized), and implementation details for each of the items.

Given the incipient state of research on the topic, no standard set of items on the topic is available for reference. The item pool was developed by researchers in the field of nuclear weapons politics and the questionnaire, which also includes questions on political attitudes and preferences, was then debriefed with experts in the field. The six themes in the item pool intend to cover general–static facts of nuclear weapons politics available to the general public that demand neither "issue expertise" nor constant media monitoring.

Data analysis procedures and software

Item selection, dimensionality analysis, and measurement invariance tests are performed using single- and multiple-group exploratory (EFA) and confirmatory (CFA) factor analytic models and unweighted least squares estimator (ULS) for categorical variables. ULS has been shown to provide more accurate standard errors compared with other estimators for categorical variables, especially for categorical variables with a small number of categories compared with other estimators for categorical variables such as the DWLS (Li, 2016; Rhemtulla et al., 2012). Fit indexes are based on the mean- and variance-adjusted chi-square (Asparouhov & Muthén, 2010), which performs best associated with the ULS estimator (Savalei & Rhemtulla, 2013). R package lavaan is used for estimation of CFA models; exploratory factor analyses are conducted using R packages semTools and psych (Jorgensen et al., 2021; R Core Team, 2020; Revelle, 2020; Rosseel, 2012). Item response function analysis is performed using R package mokken (Van der Ark, 2007).

Scale development

Items presented in Table 1 map upon different aspects of a same domain of interest, namely, general–static knowledge about nuclear weapons—their politics, development and use, therefore supporting their content validity. Per the theoretical discussion above, respondents' knowledge on nuclear weapons is hypothesized to reflect a unidimensional construct.

Nineteen items are inspected to assess their psychometric and scaling properties: (1–9) nine nuclear-armed states (the USA, Russia, China, France, the UK, North Korea, India, Pakistan, Israel), (10–12) three likely effects of the detonation of a nuclear weapon (radiation, fire, blast), (13–14) the cities bombarded with atomic weapons in World War 2 (Hiroshima, Nagasaki), (15) the death toll of the use of nuclear weapons during the Second World War, (16) the number of nuclear weapons in the world today, (17) the number of nuclear weapons in the respondent's country of residence, (18) whether

Table 1 Items measuring public knowledge about nuclear weapons, 2018

Item wording	Item format	Response options	Implementation
Some countries around the world have built nuclear weapons. Of the countries listed below, which, if any, do you believe to be currently in possession of nuclear weapons? Please select all that apply.	Multiple choice, multiple	*USA* *India* Argentina *Russia* *Pakistan* South Africa *France* *North Korea* Australia *UK* *Israel* Belarus *China* Iran Egypt None of these	Countries presented in random order; "none of these" option fixed at bottom of the list
To the best of your knowledge, how many nuclear weapons are there in the world today, approximately? If you are not sure, give your best guess.	Single response, multiple choice	< 500 1500 7500 *15,000* 50,000 100,000 or more	Response options in fixed order, ordered from lowest to highest value
How many nuclear weapons are there in your country? If you are not sure, give your best guess.	Slider, open response	0–100,000 (country specific)	Open-ended question; "I don't know" option offered
As far as you know, which of the following, if any, are the likely effects of a nuclear weapon explosion? Please select all that apply.	Multiple response, multiple choice	*Radiation* *Genetic mutations* *Fire* Erosion *A blast* Hurricanes *Loss of fertility* *Famine* None of these	Effects presented in random order; "none of these" fixed at bottom of the list.
To the best of your knowledge, has any country ever given up a nuclear weapons program? Please select one option.	Binary, close-ended	*Yes* No	Fixed order
Nuclear weapons were used during the Second World War. Identify which city or cities on which they were dropped. Please select all that apply.	Multiple response, multiple choice	*Hiroshima* *Nagasaki* Tokyo Kyoto Seoul Pyongyang I do not know	Cities presented in random order; "none of these" fixed at bottom of the list

Table 1 (continued)

Item wording	Item format	Response options	Implementation
How many people died as a result of the use of nuclear weapons during the Second World War, approximately? Please select one option.	Single response, multiple choice	< 5000 10,000–25,000 70,000–80,000 150,000–250,000 More than 500,000 I do not know	Ordered, from lowest to highest value; "I don't know" fixed at bottom of the list
To the best of your knowledge, how many nuclear-weapons tests have been conducted since 1945, approximately? If you are not sure, give your best guess.	Slider, open response	0–100,000 *(2056)*	"I don't know" option offered

Note: The correct responses for close-ended items are presented in italics in the "response options" column. The correct response for how many weapons there exist in a country are country specific; see Table A3 in the Additional file 1

any country has ever terminated its nuclear weapons program in the past, and (19) how many nuclear weapons tests have ever been carried. All items are treated as dichotomous, where incorrect responses are coded as zero and correct responses are coded as one.[4] Per the lack of previous studies on scale development for measurement of public knowledge on nuclear weaponry, exploratory procedures for item selection and assessment of dimensionality are conducted using "kitchen-sink"[5] factor analyses; item–total correlations (see Tables A1-A2 in Additional file 1) and item response function tests[6] provide auxiliary information. Items that display adequate scaling properties are retained for further analyses of dimensionality and measurement equivalence tests.

Item analysis and preliminary assessment of dimensionality

Factor analysis and item response function tests are conducted to test the adequacy of the nineteen items as observed indicators of the general–static knowledge on nuclear weapons construct. Per the assumption on unidimensionality of the construct, a one-dimension model is fitted to each of the samples.[7]

Factor loadings for seventeen items are moderate to large and averaged 0.40–0.95 across samples (see Figure A1 in the Additional file 1). Only two items (number of nuclear weapons in the respondent's country and whether any country has terminated its nuclear weapons program) severely underperform, with loadings averaging ≤ 0.20. Model fit indexes, however, provide mixed support to the one-factor solution that includes the nineteen items and disparities in model fit across countries are detected (see Table A4 in Additional file 1). Whereas the RMSEA suggests good model fit (≤ 0.06), a SRMR ≥ 0.08 is found in all samples and indicates the presence of large residual

(unexplained) correlations, violation of local independence, or even the presence of multidimensionality in the data. Residual (unexplained) correlations are examined across the eight samples, and their correspondent Cramér's V coefficient (Cramér, 1946) are calculated to evaluate local dependency (Figure A4 in the Additional file 1). Most of the 171 residual correlations and Cramér's V values in each sample are < 0.1, indicating relatively small amounts correlation unexplained by the model. Residual correlations larger than 0.15, however, are detected and might indicate violations of local independence.

Item response functions and inspection of the probability of correct response per test scores (Figure A5 in the Additional file 1) show that five items have approximately a random, fifty–fifty chance of correct answers even among respondents with the highest test scores: the casualties of the use of nuclear weapons in World War 2, the number of nuclear weapons in the world today, the number of nuclear weapons in the respondent's country of residence, whether any country has ever terminated its nuclear weapons program in the past, and how many nuclear weapons tests have ever been carried.[8] These same five items also display low discrimination and high difficulty (Figures A2-A3 in the Additional file 1.)

Put together, results from the item test score and the one-dimension factor analyses suggest that those five items are not sound indicators of general–static knowledge on nuclear weapons, or, alternatively, they might be indicators of other constructs in a multidimensional solution.

To further assess the dimensionality underlying the data, exploratory factor analyses of the nineteen items are conducted. The likelihood ratio test of multiple solutions suggests three as the optimal number of factors to be retained in seven out of eight samples.[9] The three-factor EFA indicates that the dimensions are strongly correlated, with an average correlation of 0.55 among factors. One factor accounts for the nine items on nuclear weapons possessors; a second factor accounts for the three likely effects of a nuclear weapon explosion; and a third factor accounts for the two cities bombed with atomic weapons and their associated casualties, the number of nuclear

---

[4] Per the level of difficulty of open-ended questions revealed in the responses, with a vast majority of respondents opting for the "I don't know" or delivering incorrect responses, responses coded as "correct" include the correct responses as well as a generous margin of "close enough" numbers to capture responses that might miss the exact correct answer but are fair approximations. See Table A3 in the Additional file 1.

[5] In regression analysis, the "kitchen-sink" approach refers to the practice of adding as many independent variables as possible in a model either to detect relevant predictors of the dependent variable or to increase the R^2 (Rogerson, 2001 p. 132-5); some authors refer to it as "garbage-can" approach (Achen, 2005). In the current paper, the author employs the term "kitchen-sink" to refer to exploratory data analytic procedures in which a large number of potential indicators of the hypothesized construct are tossed into the model.

[6] The item response function assesses whether the probability of correct response to a said item is associated with the test score; a steady, monotonic increase in that probability is expected among respondents who score higher in the test. Test scores are calculated stepwise, excluding the item for which the probability of correct response is being tested.

[7] See Figures A1-A3 in the Additional file 1 for item loadings, difficulty, and discrimination from the "kitchen-sink" model.

[8] It must be registered that aggregate results for the number of nuclear weapons in the respondent's country of residence are inflated by the high rate of correct responses in the Polish and Swedish samples, where about 50% of respondents delivered the correct response of 0; in the other six samples, correct responses average around 5% (see Table A3 in Additional file 1.). Among the respondents scoring the highest in the test scores, the correct response rate is around 90% in Poland and Sweden and hovers 25% on average in the other six samples. Such results indicate the presence of item bias (Van de Vijver & Leung 2011), which turns the items unsuitable for multiple-group analysis.

[9] Extraction of a fourth factor would not significantly improve the model's scaled chi-squared; see Table A5 in Additional file 1. See Table A6 in Additional file 1 for the loadings from the three-factor EFA solution rotated using the oblimin oblique rotation.

weapons in the respondent's country, and the total number of nuclear tests ever carried. That two factors represent exclusively items clustered in separate item batteries (one battery on nuclear-armed states and one battery on likely effects of a nuclear explosion), and the two strongest loadings in a third factor also come from another battery (cities bombed with nuclear weapons in World War 2), plus their strong loadings in the unidimensional solution and the strong correlation among factors might indicate that the multidimensional solution in the EFA would rather be an artifact due to testlet effects[10] resulting from questionnaire design. Finally, I subject the five items that displayed low discrimination and high difficulty in the unidimensional solution to item–total correlation analysis to assess whether they might comprise a separate dimension with internal consistency (Table A7 and Figure A6 in the Additional file 1). Results indicate low internal consistency: average item–total correlate is as low as 0.26; three items display evidence of guessing (non-trivial probability of correct answer to an item when test score is 0); and odds of correct response are barely larger than 0.50 even among those who scored the highest in the test. Altogether, the evidence presented above strongly indicates that their low performance might not due to multidimensionality in the items but rather to the item scaling properties themselves. Per the current state of the field, it is difficult to assert whether these items are overall inadequate indicators of the construct in the general public. Further research on citizens' knowledge on nuclear weapons is encouraged to use cognitive interviews to examine the poor performance of those items. Respondents may be genuinely ignorant on the issues measured by these questions, or item performances may be attributed to item or questionnaire design. Available data, however, do not permit examining those hypotheses.

On the retention and exclusion of items

Results from country-level "kitchen-sink" models that included all candidate items indicate that 10 of the items exhibit adequate scaling properties to comprise a cross-national measure of knowledge about nuclear weapons: six items on nuclear-armed states (the USA, Russia, China, India, Pakistan, Israel), the two Japanese atomic-bombed cities in World War 2, and two immediate effects of a nuclear weapon detonation (fire, a blast). Average factor loadings for those items ranged 0.55–0.85.

Nine items are not retained for the final scale. The five items discussed at the end of the previous subsection are excluded from the pool of variables retained for further analyses due to lack of solid scaling properties

as discussed above (see Table A7 and Figure A6 in the Additional file 1).[11] Three closed-ended items demonstrated weak scaling properties, namely low factor loading (< 0.4), low discrimination (< 0.3), and/or too-high difficulty (> 2) in most samples: whether any country has ever given up its nuclear weapons program, the number of existing nuclear weapons in the world, and casualties associated with the atomic bombing in World War 2.[12] The two open-ended items—the number of ever conducted nuclear weapons tests and the number of nuclear weapons in the respondent's country of residence—also proved to be items with high difficulty (> 2.5) and low discrimination (< 0.4) in six samples.

Four items with reasonable factor loadings and adequate item–test performance are excluded from the final scale as well. The factor loading for North Korea as a nuclear weapons possessor on the latent factor in the CFA is weaker than the factor loadings for other possessors (0.5 on average) and has low discrimination power (0.38). Item difficulty indicates that North Korea is recurrently among the least difficult items (− 1.25). It is hypothesized that, given the media coverage of the North Korean nuclear program in the recent past, familiarity or "having heard about it" might be scattered among respondents regardless of their overall knowledge about nuclear weapons affairs; in other words, the item may be rather measuring media consumption or awareness. This interpretation is consistent with its lower discrimination and difficulty parameters relative to items on the other nuclear weapons possessors. Although radiation presents robust factor loadings (0.6−0.8) and an overall proportion of correct answers close to 85%, being one of the easiest items in the pool, it presents lesser discriminatory power compared with other least-difficult items such as the USA or Russia as nuclear-armed states. As it will be discussed later, even a respondent at the lowest level of knowledge on nuclear weapons has about 25% chance of correctly ticking North Korea as a nuclear-armed state or radiation as one of the likely effects of a nuclear explosion.

Items on France and the UK as nuclear possessors present non-negligible item bias (Van de Vijver & Leung 2011). Although the items present acceptable scaling properties with moderate-to-strong factor loadings (> 0.6) and discrimination (> 0.5), they differ considerably in their parameter locations in, respectively, France and the UK compared to the other samples (see Figure A2 in the Additional file 1). Whereas their inclusion in

---

[10] A testlet is a cluster of items that share a common stimulus (e.g., items nested within a battery) (DeMars, 2006).

[11] See Figures A1-A3 in the Additional file 1 for item loadings, difficulty, and discrimination from the "kitchen-sink" model.

[12] The multinomial format of the items on casualties and on the number of nuclear weapons in the world might render an interesting study on the factors that led respondents to under- or overestimate the death toll of the atomic bombing and well as the size of the existing global nuclear arsenal. This analysis, however, falls beyond the scope of this paper.

Table 2 Fit indexes for confirmatory factor analysis per sample, 2018

	(A) Unidimensional (df = 35)				(B) Bifactor (df = 32)				
	χ^2	CFI	RMSEA	SRMR	χ^2	CFI	RMSEA	SRMR	ECV
United Kingdom	378.9	0.935	0.099	0.085	95.6	0.988	0.044	0.041	0.72
France	327.4	0.939	0.091	0.083	96.6	0.987	0.045	0.044	0.72
Germany	250.7	0.945	0.090	0.079	60.9	0.993	0.034	0.037	0.73
Italy	174.2	0.910	0.072	0.091	79.8	0.969	0.044	0.060	0.66
Belgium	313.6	0.882	0.103	0.104	78.2	0.980	0.044	0.050	0.63
Netherlands	311.1	0.866	0.102	0.106	71.0	0.981	0.040	0.048	0.60
Poland	160.6	0.947	0.069	0.077	52.1	0.991	0.029	0.042	0.71
Sweden	211.3	0.946	0.082	0.078	50.2	0.994	0.027	0.036	0.73

Note: All models are fitted using unweighted least squares estimator with mean- and variance-adjusted test statistics. χ^2, mean- and variance-adjusted chi-squared test statistics; *df* degrees of freedom, *CFI* comparative fit index, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean square residual, *ECV* explained common variance. For model identification, the latent factors and the latent variate underlying the observed variables have their means and variances set to 0 and to unit, respectively

single-sample studies should be considered, their inclusion in comparative studies result may bias the mean and distribution of scores. Further analysis of item performance of these four excluded variables is presented later in the text.

In summary, out of the nineteen items under consideration, ten of them displayed acceptable properties to comprise a nuclear knowledge scale: the USA, Russia, China, India, Israel, and Pakistan as nuclear weapons possessors; fire and a blast as outcomes of a nuclear weapon explosion; and Hiroshima and Nagasaki as the target of atomic bombings. These items tap on different subdomains of general—static knowledge on nuclear weapons.

Dimensionality of the proposed scale

Table 2, column A displays measures of fit for the unidimensional model including the ten selected indicators. Although the CFI for five countries (≈ 0.95) are suggestive of model acceptability, RMSEA indicates the presence of source of misfit, and the SRMR indicates the presence of large residuals. Examination of the residual correlation matrices confirms the presence of local dependence.

Cramér's V coefficient is calculated for all residual correlations in the model to evaluate local dependency (Figure A7 in the Additional file 1). Most residual correlations and Cramér's V values are < 0.1, indicating relatively small amounts of correlation left unexplained by the model. For three sets of variables—Israel, India and Pakistan as nuclear possessors; Hiroshima and Nagasaki as bombarded cities; and fire and a blast as effects of a nuclear warhead explosion—residual correlation are considerably > 0.15–0.20 and with Cramér's V usually > 0.2, indicative of moderate association. Local dependency for the three sets of items is found in all samples.

Even though the evidence suggests the model is "essentially unidimensional" (Bonifay et al., 2015), the presence

of local dependency leads to poor model fit. Importantly, ignoring local dependency can also lead to misestimation of item parameters (DeMars, 2006). An alternative approach to model the construct of interest and accommodate local dependencies is the bifactor measurement model (DeMars, 2006; Reise, 2012). A bifactor model "specifies that the covariance among a set of item responses can be accounted for by a single general factor that reflects the common variance running among all scale items and group (or specific) factors that reflect additional common variance among clusters of items, typically, with highly similar content" and assumes that the general and the group (specific) factors are all orthogonal (Reise, 2012, p. 668). The general factor represents the main construct of interest. In this analysis, local dependency is hypothesized to result from a testlet effect for items within each pair are nested within a common stimulus (i.e., within a same item battery) and for measuring a same subdomain of the construct of interest.[13] Exploratory factor analysis using bifactor rotation

---

[13] One might speculate about the absence of similar level of local dependency for other item pairs in the question on nuclear weapons possessors. I hypothesize that Israel, India, and Pakistan comprise a group of "difficult" items without forming a separate construct. The India–Pakistan doublet comprises at once the two most difficult items within that question (< 30% of respondents ticked each item) and most strongly correlated (with polychoric correlations of 0.7 or higher). Israel is endorsed by only 39% of respondents, a result that might reflect the public's perception on the country's deliberate ambiguity with regards to its nuclear weapons program (see Cohen, 2010). Finally, these are three nuclear-armed countries that do *not* hold a permanent seat at the United Nations Security Council. A model including the six possessors as indicators of the testlet/specific factor was also fit to the data. The likelihood ratio test indicates that the model with the six indicators of nuclear-armed states loading on the testlet/specific factor has a better fit to the data compared with the model with the three "difficult" items ($\Delta\chi^2 = 61.2$, $\Delta df = 5$, *sig.* < 0.01); however, the estimated loadings for the USA, Russia, and China are < |0.2|, therefore being of little substantive interest. I interpret the model improvement as due merely to the modeling of "leftover" correlations otherwise left unexplained.

with orthogonal factors[14] confirms the presence of the testlets.[15] A bifactor structure is therefore retained for further analyses.[16] The bifactor model is graphically presented in Fig. 1.

A confirmatory bifactor model is fitted to the ten items: in addition to the general factor, each of the three sets of variables presenting strong residual correlations is modeled as a specific factor, orthogonal both to the general and to the other specific factors (for the sake of statistical parsimony, factor loadings on the specific factors are constrained to equality with no detrimental impact on model fit). Table 2, column B reports measures of fit for the bifactor model for each sample. Model fit is excellent across samples: CFI $\geq$ 0.96, RMSEA < 0.05, and SRMR $\leq$ 0.06.[17] These results suggest that, in addition to a general factor that accounts for the covariance in the item pool, there are subdomains in the data represented by the specific factors that represent a share of that variance beyond what is explained by the general construct.

The rightmost column in Table 2 reports the explained common variance (ECV), which is the common variance explaiwned by the general factor divided by the total common variance. This ratio assesses the relative strength of the general factor and has been described as a coefficient of "closeness to unidimensionality" (Ten Berge & Sočan, 2004, p. 621; see also Rodriguez et al., 2016). ECV values are 0.6–0.7, meaning that approximately 60 to 70% of the common variance is explained by the general factors.

An auxiliary index, the Percentage of Uncontaminated Correlations (PUC; Bonifay et al., 2015), assesses the ratio of unique correlations in a test attributed to the general factor only relative to all correlations in a test—i.e., the correlations "uncontaminated" by specific factors or testlets. Only 5 out of [(10 items $\times$ (10 items $-$ 1))/2] = 45 correlations between pairs of variables map upon specific/testlet factors, meaning that 40 correlations inform on the general factor only and results in a PUC of (40/45) $\approx$ 0.89.[18]

The results are supportive of the claim that the items in the test tap on the target trait they were designed to measure. Nine out of ten unique correlations map upon the general factor only, indicating that it accounts for the lion's share of all common variance among the items. The ECV nevertheless indicates that the testlet/specific factors account about 30% of the common variance; as discussed above, the covariance not explained by the general factor might be due to questionnaire design that lead to the presence of testlets. The bifactor model accounts for the presence of testlets and has a superior model fit compared with the unidimensional solution. A comparison between the bifactor solution and unidimensional alternatives for the computation of scores is discussed below in the section on recommendations for the use of the scale.

Measurement equivalence

For valid cross-national comparisons, it is necessary to first establish the invariance of model parameters across subpopulations to warrant equivalence of the data-generating processes between them. The invariance of measurement parameters of the proposed bifactor model is assessed using multiple-group confirmatory factor analysis (Avvisati et al., 2019; Davidov et al., 2014; Jöreskog, 1971; Meredith, 1993; Vandenberg & Lance, 2000).[19] Starting with the configural model, which tests whether a same factor structure fits the data from all groups, consecutive constraints are imposed on the model to test for the invariance of thresholds, the invariance of loadings (metric invariance), and item intercepts (scalar invariance); invariance of unique variances (strict invariance) may also

---

[14] Exploratory bifactor analysis is an exploratory factor analytic model with a rotation criterion that allows all items to freely load on the first factor (which represents the general factor) and encourages a perfect cluster structure for the loadings on the other factors (Jennrich & Bentler, 2011).

[15] One additional US–Russia testlet factor emerged in the bifactor EFA for three samples only. The proportion of explained variance attributed to it is very small ($\leq$ 0.07) in the three samples. Finally, the Cramér's V associated to the item pair is weak ($\leq$ 0.13) in all samples. The testlet is treated as a nuisance and not modeled.

[16] As an additional test of whether the bifactor or the three-factor model should be preferred, a three-dimensional confirmatory factor model is fit to the data where each dimension item battery (possessors of nuclear weapons, effects, cities bombed with atomic weapons) corresponds to one dimension; see Table A9 in Additional file 1 for model fit in each sample. The bifactor and the three-factor models are compared using the likelihood ratio test (Table A10 in Additional file 1). The bifactor model outperformed the three-factor solution in all samples.

[17] Examination of residual correlations show dramatic reduction of local dependency, with residual correlations rarely exceeding 0.1. Three residual correlations notoriously > |0.15| involve India and Nagasaki in Belgium (0.165) and the Netherlands (0.157). These correlations tap on items from different questions and are not the result of a testlet effect. They are noted but not further modeled.

[18] Coefficients omega (ω) and omega hierarchical (ω_{H}) provide additional support to the adequacy of the bifactor models (Rodriguez et al., 2016). Omega hierarchical shows that 81−88% of the total variance of unit-weighted composites could be attributed to the general factor. Omega indicates that the bifactor model accounts for 91−97% of the total variance of total scores; in other words, the specific factors account for only approximately 10% of the variance in total scores (see Table A8 in the Additional file 1).

[19] A detailed treatment of measurement equivalence is beyond the scope of this work. See Millsap and Yun-Tein (2004) and Wu and Estabrook (2016) for a discussion on measurement invariance in the context of factor analysis for categorical manifest variables.
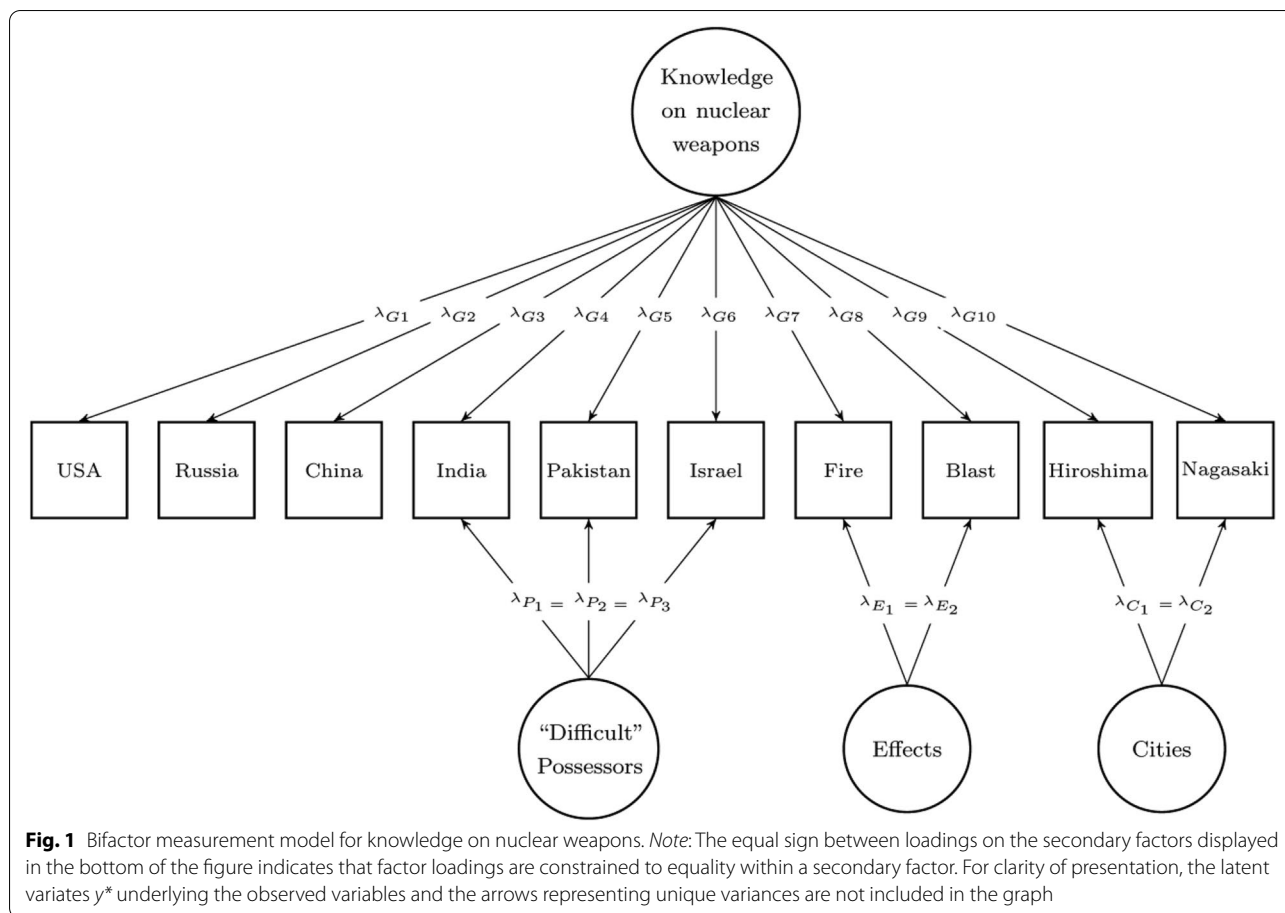
Fig. 1 Bifactor measurement model for knowledge on nuclear weapons. *Note*: The equal sign between loadings on the secondary factors displayed in the bottom of the figure indicates that factor loadings are constrained to equality within a secondary factor. For clarity of presentation, the latent variates *y** underlying the observed variables and the arrows representing unique variances are not included in the graph

be tested.[20] More restrictive models might display loss of fit compared with lesser restrictive models; if deterioration in model fit is nevertheless small, it should not be interpreted as lack of invariance. Recommendations for the use of fit indexes to test for measurement invariance in congeneric measurement models with continuous indicators such as ΔCFI smaller than or equal to -0.01 and ΔRMSEA smaller than or equal to $+0.015$ supplemented by ΔSRMR smaller than or equal to $+0.03$ for invariance of loadings and ΔSRMR smaller

than or equal to -0.01 for invariance of intercepts (Chen, 2007; see also Cheung & Rensvold, 2002) have been documented in the literature, but lesser progress has been made for bifactor models and for models with categorical indicators. Khojasteh and Lo (2015) suggested ΔCFI approximately equal to -0.004 for metric invariance in bifactor models, but no recommendation is made for scalar or strict invariance. Moreover, most simulation-based recommendations are based on two-group models, and it is admissible that minor deviance accumulated across a large number of groups might lead to the rejection of an otherwise acceptable invariant model. To test for different levels of measurement, examination of ΔCFI, ΔRMSEA, and ΔSRMR is complemented by the expected parameter change (EPC; Oberski et al., 2015) for parameters constrained to equality.

A caveat is nevertheless in order before proceeding to the measurement invariance test. Model identification for congeneric single- and multiple-group factor analytic models for dichotomous outcomes has been well established in the literature (e.g., Christofferson, 1975; Muthén, 1984; Wu & Estabrook, 2016). Model

---

[20] Model identification conditions have been established for measurement invariance in congeneric models with binary indicators (Wu & Estabrook, 2016) but not for bifactor models. I tentatively apply Wu and Estabrook's recommendation for congeneric models to a bifactor model: for the inadequacy of testing invariance of thresholds and of loadings separately for binary items, once configural equivalence is established, I first test the invariance of thresholds and loading simultaneously followed by invariance of item intercepts, with between-group equality parameter constraints imposed simultaneously to the release of unnecessary identification constraints to test a said level of equivalence (Wu & Estabrook, 2016). Given the complexity of the bifactor model and because the test of invariance of unique variance requires the use of theta parameterization, which may be numerically unstable under certain circumstances (Wu & Estabrook, 2016), invariance of unique variance requires will not be tested.

Table 3 Fit indexes for invariance tests, 2018

	χ^2	df	CFI	RMSEA	SRMR
Configural	573.7	256	0.987	0.039	0.049
Panel A: measurement invariance, latent variate scales unconstrained (except in the reference group)					
A1. Thresholds and loadings	569.7	249	0.987	0.040	0.049
A2. + Intercepts	662.5	291	0.985	0.039	0.050
Panel B: measurement invariance, latent variate scales fixed across groups					
B1. Thresholds and loadings	568.5	319	0.990	0.031	0.057
B2. + Intercepts	676.1	361	0.987	0.033	0.057
Panel C: structural invariance, latent variate scales fixed across groups					
C1. Variances: general and specific, constrained to unit	1012.5	389	0.975	0.044	0.080
C2. Variances: general released, specific constrained to unit	724.1	382	0.986	0.033	0.059
C3. + Means: general and specific constrained to zero	1702.7	410	0.948	0.062	0.059
C4. + Means: general factor released, specific fixed to zero	1263.7	403	0.965	0.051	0.059

Note: All models are estimated using unweighted least squares estimator with mean- and variance-adjusted test statistics. Models use delta parameterization.
χ^2 chi-squared test statistics, *df* degrees of freedom, *CFI* comparative fit index, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean square residual

identification and measurement invariance procedures for bifactor model variables remain nevertheless understudied. Wu and Estabrook (2016) demonstrate that invariance of thresholds for polytomous categorical variables—which have two or more thresholds—equates the scales of the latent responses y^* underlying the observed categorical variables y and therefore allows for metric and scalar invariance tests and the comparison of latent variances and means; imposing invariance on a second threshold in dichotomous variables is impossible. Moreover, in bifactor models, item-shared variances are also caused by secondary factors, and rules of identification might differ from those for congeneric models. To tentatively address some of those issues, in special, the scaling of the latent variates, two sets of results will be presented in Table 3: panel A reports results invariance tests for models with unconstrained latent variate scales (except for the reference group);[21] panel B reports results with latent variate scales constrained to equality between groups.[22] Further research on the topic is encouraged.

The excellent fit indexes displayed in the top row of Table 3 indicate that configural invariance holds in the data. The successive imposition of equality constraints to

test for the invariance of thresholds and loadings (Model A1) as well as of intercepts (Model A2) do not deteriorate the model fit and support measurement equivalence across samples: ΔCFI, ΔRMSEA, and ΔSRMR for across all levels of invariance are never > 0.02. Measurement invariance is also supported by the EPC test. Measurement invariance is established for the ten-item scale with freely estimated latent variate scales. No Heywood case was detected. Likewise, results from Table 3, panel B indicate that measurement invariance is held with the scale of latent variates constrained to unit as well (Model B1–B2). ΔSRMR ≈ +0.01 in panel B models compared with panel A suggests a minor increment in average unexplained correlations which might be otherwise accounted by varying latent variate scales. Such a fit deterioration is nevertheless small and models in panel B are not rejected. The (scaled) likelihood ratio test indicates that models with fixed and freed scales of y^* are equivalent, and the additional constraints on the scale of y^* do not deteriorate model fit (invariance of thresholds and loadings: $\Delta\chi^2 = 61.7$, $\Delta df = 70$, *sig.* = 0.75; invariance of intercepts: $\Delta\chi^2 = 70.3$, $\Delta df = 70$, *sig.* = 0.47).

Results from Table 3, panel C indicate the presence of structural non-invariance across countries. Between-samples equality of variance of the general and the specific factors (Model C1) is rejected. Even though the deterioration of the fit indexes is modest, the EPC test rejects the between-group equality constraints imposed on the general factor. Once the general factor variance is released to vary between groups, equal variance of the secondary factors (Model C2) cannot be rejected as fit index deterioration is minimal. Holding the variance of the specific factors constrained to equality across

---

[21] For identification purposes, the scale of latent variate y^* is fixed to unit in single-group and multiple-group configural models; after thresholds and loadings are constrained to equality, the scale of y^* and the variance of the latent factor remain fixed in the reference group only (Wu & Estabrook, 2016).

[22] In these models, the scale of all latent variates y^* are parsimoniously fixed to unit for all variables in all samples. Examination of the scale parameters for models in Table 3, panel A suggests that for the virtually all indicators in the eight samples have the scale of their latent variates y^* close to unit and with their confidence intervals including it.

Table 4 Loadings and thresholds for measurement invariant model, 2018

	Unidimensional		Bifactor				
	General	Threshold	General	Specific 1	Specific 2	Specific 3	Threshold
United States	0.90	− 0.97	0.94				− 0.93
Russia	0.88	− 0.93	0.92				− 0.89
China	0.79	− 0.41	0.82				− 0.37
Israel	0.74	0.11	0.67	0.51			0.07
India	0.84	0.47	0.78	0.51			0.43
Pakistan	0.72	0.38	0.65	0.51			0.34
Fire	0.60	− 0.34	0.56		0.62		− 0.58
Blast	0.65	− 0.58	0.62		0.62		− 0.82
Hiroshima	0.78	− 1.03	0.75			0.57	− 0.85
Nagasaki	0.76	− 0.47	0.73			0.57	− 0.28

Note: Parameters for models with invariant thresholds, loadings, and intercepts, with the scale of latent variates $y*$ fixed to unit. Right-side panel reports parameters for the bifactor model in Table 3, model B2. Unidimensional model with invariant parameters included in the left-side. All parameters are statistically significant at $p > 0.01$

samples, constraints to equalize latent means for the general and for the specific factors across countries are also rejected (Model C3–C4). Such results indicate country-level differences in the location (mean) and distribution (variance) of the public's knowledge on nuclear weapons.

Loadings and thresholds for bifactor model C2 are presented in Table 4. All item intercepts are fixed to zero.[23] For the sake of comparison with its unidimensional counterpart (with invariant thresholds, loadings, and intercepts; scale of latent variate $y*$ fixed to unit; CFI = 0.92; RMSEA = 0.075; SRMR = 0.1), loadings and thresholds for the invariant unidimensional model are also reported. Factor loadings for both models are very similar in magnitude, which suggests that the inclusion of specific factors in the bifactor model does not "steal" explained common variance from the general factor and, importantly, genuinely account for variation left unexplained by the general factor. In other words, the specific factors are not methodological artifacts.

A note on four excluded items: low discrimination and item bias

Four items in the initial pool were not retained for the dimensionality and measurement invariance analysis despite of strong face value: radiation as effect of a nuclear explosion; and North Korea, France, and the UK as possessors of nuclear weapons. Exploratory data analysis using item–total correlations and "kitchen-sink" factor analytic models suggested that those items present either low discrimination (radiation, North Korea) or item bias (France, the UK). Given the novelty of the
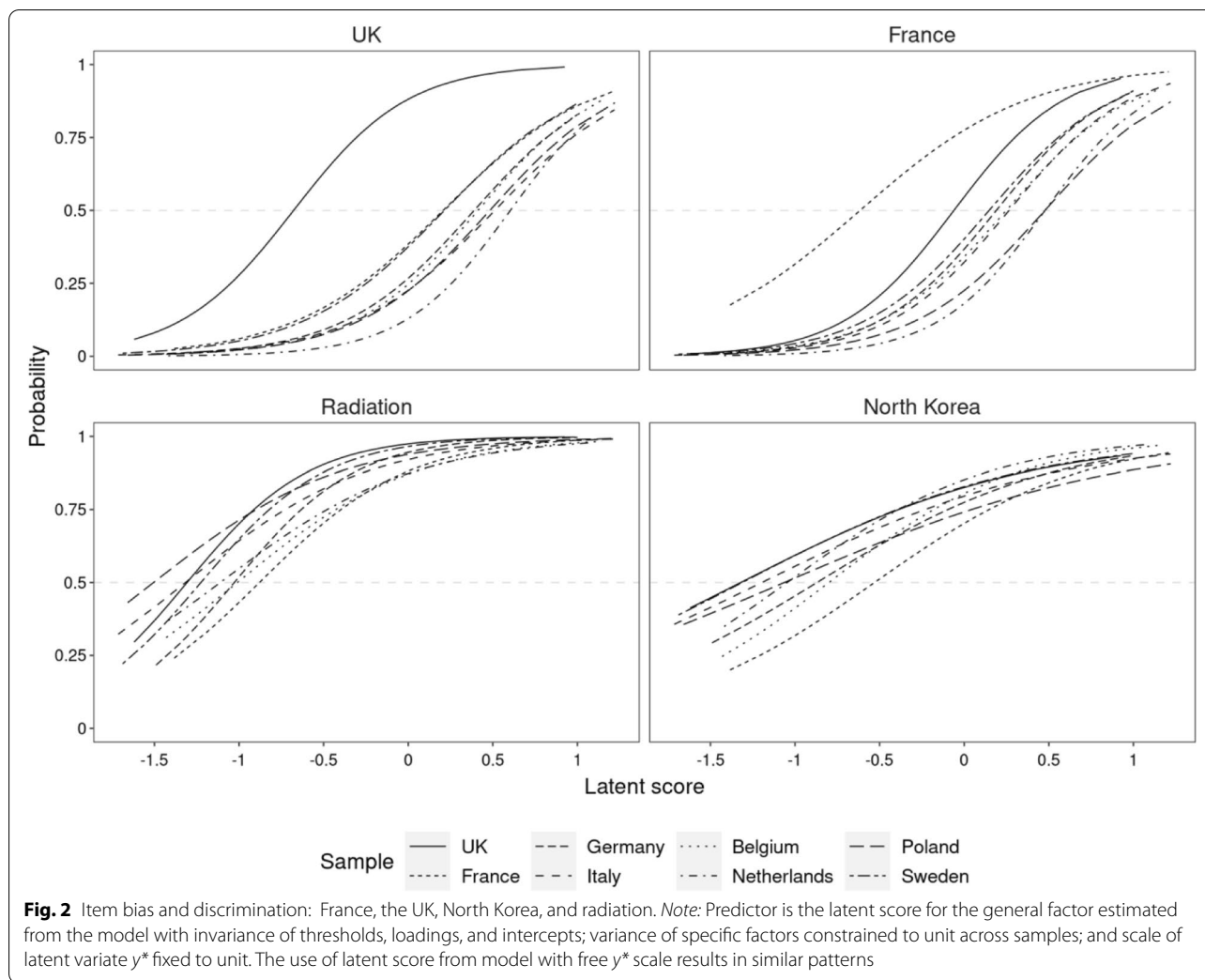
instrument, the non-inclusion of such items should nevertheless be further justified. A reassessment of such items is performed after the invariance of the ten-item scale has been established. The factor scores for each sample are estimated (from Table 3, Model C2), and the four excluded items are regressed on the general factor scores; predicted probabilities are presented in Fig. 2.

Predicted probabilities for the UK and France indicate the presence of item bias. Endorsement of each of the two items is easier in their "home samples", as indicated by the country's probability curve locations; in the case of the France item in the French sample, the lower asymptote also suggests a high probability of item endorsement by chance. Predicted probabilities for radiation and for North Korea confirm the low discrimination power for both items, and the lower asymptote indicates a relatively high level (ranging from 0.25 to 0.50) of correct response due to guessing. Therefore, the retention of the former two items might lead to bias in parameter estimates (for

Table 5 Fit indexes for bifactor confirmatory factor analysis per sample, 2019

	Bifactor (df = 32)				
	χ^2	CFI	RMSEA	SRMR	ECV
United Kingdom	83.5	0.990	0.040	0.038	0.73
France	51.1	0.992	0.024	0.038	0.64
Germany	61.7	0.987	0.036	0.048	0.72
Italy	50.2	0.990	0.028	0.044	0.69
Belgium	42.9	0.994	0.022	0.040	0.67
Netherlands	39.9	0.996	0.019	0.036	0.65
Poland	52.5	0.992	0.031	0.042	0.76
Sweden	89.3	0.977	0.051	0.056	0.68

---

[23] A full list of estimated parameters for Model C2, Table 3 are reported in the Additional file 2.

Fig. 2 Item bias and discrimination: France, the UK, North Korea, and radiation. *Note:* Predictor is the latent score for the general factor estimated from the model with invariance of thresholds, loadings, and intercepts; variance of specific factors constrained to unit across samples; and scale of latent variate *y** fixed to unit. The use of latent score from model with free *y** scale results in similar patterns

instance, inflation of latent means in France and the UK), and the retention of the latter two items would not contribute to discrimination among respondents' abilities.

Replication and validation

Data from the 2019 survey is used to validate the model proposed above. Table 5 reports the fit indexes and ECV for the bifactor model in the eight countries in the validation sample.

Results presented in Table 5 indicate excellent model fit also in the validation sample: CFI $\geq$ 0.98, RMSEA < 0.05, and SRMR $\leq$ 0.06. A comparison between Tables 2 and 5 shows highly similar model fit in the two surveys. As in the 2018 sample, ECV demonstrates that the general factor accounts for approximately two-thirds of the common variance in the data.

Regarding measurement invariance, once again, results for the validation sample largely resemble those for the calibration sample. Configural invariance holds in the data, as

shown by the excellent fit indexes in the top row of Table 6. Invariance of factor loadings and thresholds as well as invariance of intercepts are held with the scale of the latent variate *y** either allowed to vary across groups (Table 6, panel A) or fixed to unit in all groups (Table 6, panel B).

Results from Table 6, panel C indicate a lack of structural invariance across countries also in the validation sample. Having in mind the differences in demographic composition of the 2018 and 2019 samples, there is some evidence supporting invariance of latent variances in the 2019 validation sample (Model C1): deterioration of the fit indexes is modest, and the EPC test does not provide strong grounds to reject invariance of latent variances. Moreover, releasing the latent variance of the general factor to be freely estimated across samples (Model C2) slightly decreases model fit except for the SRMR (which means that Model C2 leaves smaller unexplained correlations compared with Model C1). The likelihood ratio test shows that the models have equivalent performances

Table 6 Fit indexes for invariance tests, 2019

	χ^2	df	CFI	RMSEA	SRMR
Configural	458.8	256	0.990	0.032	0.046
Panel A: measurement invariance, latent variate scales unconstrained (except in the reference group)					
A1. Thresholds and loadings	480.1	249	0.989	0.035	0.048
A2. + Intercepts	653.8	291	0.982	0.040	0.050
Panel B: measurement invariance, latent variate scales fixed across groups					
B1. Thresholds and loadings	733.6	319	0.979	0.041	0.071
B2. + Intercepts	883.3	361	0.974	0.043	0.071
Panel C: structural invariance, latent variate scales fixed across groups					
C1. Variances: all constrained to unit	898.0	389	0.975	0.041	0.079
C2. Variances: general released, specific constrained to unit	921.6	382	0.973	0.043	0.072
C3. Means: general and specific constrained to zero	1456.9	410	0.948	0.057	0.072
C4. Means: general factor released, specific fixed to zero	1367.5	403	0.952	0.055	0.072

Note: All models are estimated using unweighted least squares estimator with mean and variance adjusted test statistics. Models use delta parameterization. χ^2, chi-squared test statistics; *df*, degrees of freedom; *CFI*, comparative fit index; *RMSEA*, root mean square error of approximation; *SRMR*, standardized root mean square residual.

($\Delta\chi^2 = 9.1$, Δdf $= 7$, *sig.* $= 0.25$). To continue the comparisons between the calibration and the validation samples, Model C2 is retained for the test of equality restriction on the latent means. Similar to the 2018 calibration sample, equality of the latent means in the general factor as well as equality of the latent means in the specific factors are rejected.[24]

Evidence from the validation analysis displays excellent model fit in each of the eight countries and supports measurement invariance as well. Similar results are also found for the tests of invariance of latent variances and means in the calibration and validation samples. Taken together, results presented in Tables 5 and 6 provide strong cross-national evidence of validation for the model.

Correlation with criterion variables

Evidence presented above demonstrates the scaling and measurement invariance properties of the proposed measurement model of public knowledge about nuclear weapons. Next, it is discussed whether the construct correlates with other variables presumed to be part of its nomothetic span (Embretson, 1983). Covariance with two criterion variables, one demographic and one attitudinal, are assessed: education (five points, from lesser than complete elementary education to higher education) and the perceptions that nuclear weapons testing caused environmental damage (four points, from strongly disagree to strongly agree). Education is one of the major predictors of political behavior and political information

(Delli Carpini & Keeter, 1996) and therefore is expected to be also positively correlated to knowledge on nuclear weapons. The environmental damage caused by nuclear weapons testing has been documented (Beck et al., 2010, Prăvălie, 2014) and, importantly, discussed in popular media outlets (ABC News, 2017, Rust, 2019, Welt Documentary, 2020). It is expected that the more knowledgeable a citizen is about nuclear weapons politics, the more one will be aware of the environmental impact of nuclear weapons testing. The correlations between the construct of interest and the criterion variables are computed using structural equation models including one criterion a time, where the criterion variable correlates with the general factor only.

Results in Table 7 show the correlation of the general factor with educational achievement and with perceptions of environmental damage estimated using the 2019 survey. The average correlation between the general factor and education is approximately 0.20, ranging from 0.14 to 0.30. Education seems not to be the only predictor of knowledge about nuclear weapons but is an important one nevertheless; results in Table 7 are similar to correlations between education and political information and political thinking found in previous studies ($r = 0.28$ in Neuman (1981); first difference $= 0.27$ in Barabas et al. (2014); unstandardized regression coefficient $= 016–0.37$ in Zaller (1986)). Correlations with perceptions of environmental damage caused by nuclear weapons testing average 0.30 across samples, ranging from 0.15 in the United Kingdom to 0.47 in Germany; in five out of eight samples, correlations are 0.3 or higher, providing strong evidence that knowledge on nuclear weapons may influence perceptions and preferences on the matter.

---

[24] A full list of estimated parameters for Model C2, Table 6 are reported in the Additional file 3.

Table 7 Correlation between the general factor in the bifactor model and knowledge on nuclear weapons and criterion variables, 2019

	Education	Nuclear tests caused environmental damage
United Kingdom	0.165	0.148
France	0.192	0.319
Germany	0.304	0.474
Italy	0.148	0.325
Belgium	0.122	0.373
Netherlands	0.144	0.343
Poland	0.193	0.256
Sweden	0.234	0.172

Note: All correlation are statistically significant at $p \leq 0.01$

Guidelines on the use of the scale

The discussion above shows that the proposed bifactor measurement model for the assessment of the public's knowledge on nuclear weapons displays solid psychometric properties and outperforms alternative unidimensional solutions due to the presence of testlets; the specific factors in the bifactor model, in addition to the general factor representing the construct of interest, account for testlet effects caused by questionnaire design. Per the structural complexity of the bifactor model, the implications of using unidimensional representations of the construct—such as summated scores—in applied research deserves consideration.

Scores are estimated using five different approaches: (1) latent factor scores from the bifactor model (scores from the general factor are of main interest); (2) latent factor scores from the unidimensional model; (3) a unit-weighted sum of the item; (4) a weighted sum of items using the factor loadings on the general factor in the bifactor model as weights; and (5) a weighted sum of items using the factor loadings on the factor in the unidimensional model as weights. The five obtained scores are *highly* correlated with each other, $r \geq 0.96$ (Table A11 in the Additional file 1), indicating that factor scores from different solutions or summated scores will order observations in a virtually identical manner.

Polyserial correlations between the five scores and the same criterion variables used in Table 7 are computed and reported in Table 8. The first striking result is the similarity of coefficients obtained regardless of which of the five scores is used. However, the most important finding comes from a comparison between correlation coefficients reported in Tables 7 and 8. Correlations reported in Table 8 are systematically lower compared with those in Table 7. In some cases, the polyserial correlations—e.g., of nuclear knowledge with education

in Italy or with perception of environmental damage in Sweden—are about one-third lower using test scores relative to correlation coefficients obtained via structural equation models. The correlations may be attenuated due to the presence of measurement error in correlation and regression analysis, whereas structural equation modeling has the measurement model embedded and therefore accounts for measurement error in the estimation of parameters. Additionally, ignoring the presence of local dependency—modeled by the specific/testlet factors—may lead to misestimation of model parameters (DeMars, 2006).[25]

Therefore, whenever possible, it is strongly recommended for applied researchers to favor the bifactor measurement instead of the unidimensional solution and to estimate parameters of interest within a structural equation model framework. In situations where the use of structural equation model is not a feasible option, researchers should mind that the obtained estimates may be attenuated or biased.

Discussion

Extensive effort has been dedicated to the study of attitudes toward nuclear weapons. Lesser attention has been devoted to what individuals know about nuclear weapons. This paper aims to contribute to fill this gap advancing a measurement model of nuclear weapons knowledge capable to summarize information from multiple indicators that taps on general–static, "structural" aspects of nuclear weapons history and politics rather than on knowledge or familiarity with the salient issues of the day and that permits the examination of individual and group-level differences as well as the association of knowledge with other variables of interest. To the best of the author's knowledge, this is the first systematic effort to construct and validate a measure of knowledge about nuclear weapons in the general public. We believe this is an important initial step for the measurement of the public's knowledge on such an important topic in international politics.

A bifactor model with a strong general factor representing the construct of substantive interest outperforms alternative solutions and is supported by data from eight European countries. The presence of testlet factors in the latent structure is noted, but it is demonstrated that the

---

[25] Ignoring local dependency may have consequences for the estimation of model parameter within the structural equation model framework as well. Table A12 in Additional file 1 reports the correlations between the common factor and the criterion variables in a unidimensional factor analytic solution, which accounts for measurement error in the observed variables but does not account for the presence of the specific/testlet factors. Correlations in Table A12 are attenuated in comparison to correlations in Table 7.

Table 8 Correlation between test scores and criterion variables, 2019

		(1)	(2)	(3)	(4)	(5)
United Kingdom	Education	0.133	0.144	0.143	0.142	0.143
	Envir. damage	0.101	0.096	0.114	0.116	0.113
France	Education	0.153	0.152	0.151	0.154	0.154
	Envir. damage	0.245	0.238	0.250	0.251	0.247
Germany	Education	0.252	0.266	0.261	0.262	0.264
	Envir. damage	0.378	0.378	0.376	0.377	0.376
Italy	Education	0.110	0.102	0.104	0.109	0.108
	Envir. damage	0.244	0.24	0.244	0.246	0.244
Belgium	Education	0.090	0.100	0.098	0.098	0.099
	Envir. damage	0.285	0.271	0.284	0.284	0.281
Netherlands	Education	0.123	0.141	0.146	0.136	0.141
	Envir. damage	0.276	0.277	0.277	0.276	0.275
Poland	Education	0.166	0.175	0.188	0.178	0.180
	Envir. damage	0.194	0.184	0.194	0.201	0.198
Sweden	Education	0.199	0.212	0.204	0.203	0.205
	Envir. damage	0.120	0.102	0.127	0.125	0.120

Note: The scores

(1) Latent factor score, general factor, bifactor model.

(2) Latent factor score, unidimensional model.

(3) Summated score, unit weight.

(4) Summated score, weighed by loadings on the general factor, bifactor model.

(5): Summated score, weighed by loadings, unidimensional model

general factor accounts for the lion's share of common variance among the observed variables. Measurement invariance across eight samples has been established, which indicates that the construct can be meaningfully compared across contexts. Moreover, the construct of interest correlates as expected with demographic and attitudinal criterion variables. Guidelines on the operationalization of the scale in future studies are also provided.

Cross-national differences in latent means and variance of the constructs are reflected in the lack of structural invariance and deserve further investigation. Although the explanation of those cross-national differences falls beyond the scope of this article, we hypothesize that the public's knowledge on the nuclear weapons can be at least in part due to media structure and its role in diffusion of information, and to the impacts of social inequalities on information access (Curran et al., 2009; Grönlund & Milner, 2006).

Per the novelty of the topic and the absence of other widely replicated (and validated) measures of knowledge about nuclear weapons, we encourage researchers working on public opinion and international security to replicate and eventually refine and update the measurement model proposed in this article. Researchers are also invited to further expand the array of questions and subdomains on knowledge about nuclear weapons being measured and to assess how measures of knowledge on general and static aspects of nuclear weapons politics do correlate with awareness to "breaking news" and policy issues on the matter. This study aims to be a first step toward a more encompassing understanding of what the general public does and does not know about nuclear weapons.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42409-021-00028-5.

Additional file 1:. Supplementary material.

Additional file 2:. Estimated parameters, Table 3, Model C2.

Additional file 3:. Estimated parameters, Table 6, Model C2.

References

ABC News. 2017. This concrete dome holds a leaking toxic timebomb [Video]. YouTube. https://www.youtube.com/watch?v=autMHvj3exA

Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, *22*(4), 327–339.

Asparouhov, T. & Muthén, B. (2010). Simple second order chi-square correction. Mplus techncal appendix.

Avvisati, F., Le Donné, N., & Paccagnella, M. (2019). Cross-cultural comparability of questionnaire measures in large-scale international surveys. *Measurement Instruments for the Social Sciences*, *1*(8), 1–10.

Barabas, J., Jerit, J., Pollock, W., & Rainey, C. (2014). The question(s) of political knowledge. *American Political Science Review*, *108*(4), 840–855.

Beck, H. L., Bouville, A., Moroz, B. E., & Simon, S. L. (2010). Fallout deposition in the Marshall Islands from Bikini and Enewetak nuclear weapons tests. *Health Physics*, *99*(2), 124–142.

Boehnke, K., Macpherson, M. J., Meador, M., & Petri, H. (1989). How West German adolescents experience the nuclear threat. *Political Psychology*, *10*(3), 419–443.

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT Multidimensionality Index. *Structural Equation Modeling*, *22*(4), 504–516.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255.

Christoffersson, A. (1975). Factor analysis of dichotomized variables. *Psychometrika*, *40*(1), 5–32.

Cohen, A. (2010). *The worst-kept secret: Israel's bargain with the bomb*. New York: Columbia University Press.

Cortright, D., & Mattoo, A. (1996). Elite public opinion and nuclear weapons policy in India. *Asian Survey*, *36*(6), 545–560.

Cramér, H. (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Curran, J., Iyengar, S., Lund, A. B., & Salovaara-Moring, I. (2009). Media system, public knowledge and democracy: A comparative study. *European Journal of Communication*, *24*(1), 5–26.

Dahl, R. A. (1985). *Controlling nuclear weapons: Democracy versus guardianship*. Syracuse: Syracuse University Press.

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

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. New Haven: Yale University Press.

DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement*, *43*(2), 145–168.

Egeland, K., & Pelopidas, B. (2021). European nuclear weapons? Zombie debates and nuclear realities. *European Security*, *30*(2), 237–258.

Eichenberg, R. C. (1998). Domestic preferences and foreign policy: Cumulation and confirmation in the study of public opinion. *Mershon International Studies Review*, *42*(1), 97–105.

Ellsberg, D. (2017). *The doomsday machine: Confessions of a nuclear war planner*. New York: Bloomsbury.

Embretson, S. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*(1), 179–197.

Fiske, S. T., Pratto, F., & Pavelchak, M. A. (1983). Citizens' images of nuclear war: Content and consequences. *Journal of Social Issues*, *39*(1), 41–65.

Flynn, G., & Rattinger, H. (1985). The public and Atlantic defense. In: *The public and Atlantic defense*, ed G Flynn & Rattinger H. London: Routledge.

Graham, T. W. (1988). The pattern and importance of public knowledge in the nuclear age. *Journal of Conflict Resolution*, *32*(2), 319–334.

Grönlund, K., & Milner, H. (2006). The determinants of political knowledge in comparative perspective. *Scandinavian Political Studies*, *29*(4), 386–406.

Haste, H. (1989). Everybody's scared--but life goes on: Coping, defense and action in the face of nuclear threat. *Journal of Adolescence*, *12*(1), 11–26.

Haworth, A. R., Sagan, S. D., & Valentino, B. A. (2019). What do Americans really think about conflict with nuclear North Korea? The answer is both reassuring and disturbing. Bulletin of the Atomic Scientists, 75(4), 179–186.

Herron, K. G., & Jenkins-Smith, H. C. (2006). *Critical masses and critical choices: Evolving public opinion on nuclear weapons, terrorism, and security*. Pittsburgh: University of Pittsburgh Press.

Herron, K. G., & Jenkins-Smith, H. C. (2014). Public perspectives on nuclear security. *Risk, Hazards & Crisis in Public Policy*, *5*(2), 109–133.

Herzog, S., & Baron, J. (2017). Public support, political polarization, and the nuclear-test ban: Evidence from a new US National Survey. *Nonproliferation Review*, *24*(3-4), 357–371.

Iyengar, S. (1990). Shortcuts to political knowledge: The role of selective attention and accessibility. In J. A. Ferejohn, & J. H. Kuklinski (Eds.), *Information and democratic processes*. Urbana: University of Illinois Press.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*(4), 537–549.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*(4), 409–426.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2021). "semTools: Useful tools for structural equation modeling." R package version 0.5-4.

Khojasteh, J., & Lo, W.-J. (2015). Investigating the sensitivity of goodness-of-fit indices to detect measurement invariance in a bifactor model. *Structural Equation Modeling*, *22*(4), 531–541.

Knopf, J. W. (2012). The concept of nuclear learning. *Nonproliferation Review*, *19*(1), 79–93.

Kramer, B. M., Michael Kalick, S., & Milburn, M. A. (1983). Attitudes toward nuclear weapons and nuclear war: 1945-1982. *Journal of Social Issues*, *39*(1), 7–24.

Kristensen, H. M. (2014). Nuclear weapons modernization: a threat to the NPT? *Arms Control Today*, *44*(4), 8–15.

Krosnick, J. A. (1990). Government policy and citizen passion: A study on issue publics in Contemporary America. *Political Behavior*, *12*(1), 59–92.

Li, C.-H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387.

McAllister, I., & Mughan, A. (1986). The nuclear weapons issue in the 1983 British general election. *European Journal of Political Research*, *14*(5-6), 651–667.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.

Millsap, R. E., & Yun-Tein, J. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research*, *39*(3), 479–515.

Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*(1), 115–132.

Neuman, W. R. (1981). Differentiation and integration: Two dimensions of political thinking. *American Journal of Sociology*, *86*(6), 1236–1268.

Oberski, D. L., Vermunt, J. K., & Moors, G. B. D. (2015). Evaluating measurement invariance in categorical data latent variable models with the EPC-interest. *Political Analysis*, *23*(4), 550–563.

Pelopidas, B. (2017). *The next generation of European citizens facing nuclear weapons: Forgetful, indifferent but supportive? EU Non-proliferation Paper no. 56*. Stockholm: EU Non-Proliferation Consortium.

Pierce, J. C., Lovrich, N. P., & Dalton, R. J. (2000). Contextual influences on environmental knowledge: Public familiarity with technical terms in nuclear weapons production in Russia and the United States. *Environment and Behavior*, *32*(2), 188–208.

Prăvălie, R. (2014). Nuclear weapons tests and environmental consequences: A global perspective. *Ambio*, *43*(6), 729–744.

Press, D., Sagan, S. D., & Valentino, B. A. (2013). Atomic aversion: Experimental evidence on taboos, traditions, and the non-use of nuclear weapons. *American Political Science Review*, *107*(1), 188–206.

R Core Team (2020). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing https://www.R-project.org/.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696.

Revelle, W. (2020). *psych: Procedures for psychological, psychometric, and personality research*. R package version 2.0.8.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150.

Rogerson, P. A. (2001). *Statistical methods for geography*. London: Sage.

Rosenbaum, R. (2011). *How the end begins: The road to a nuclear World War III*. London: Simon & Schuster.

Rosi, E. J. (1965). Mass and attentive opinion on nuclear weapons tests and fallout, 1954-1963. *Public Opinion Quarterly*, *29*(2), 280–297.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Russett, B. (1990-1991). Doves, hawks, and U.S. public opinion. *Political Science Quarterly*, *105*(4), 515–538.

Russett, B., & Deluca, D. R. (1983). Theater nuclear forces: Public opinion in Western Europe. *Political Science Quarterly*, *98*(2), 179–196.

Rust, S. (2019). How the U.S. betrayed the Marshall Islands, kindling the next nuclear disaster. *Los Angeles Times*, 10 November 2019. Available at https://www.latimes.com/projects/marshall-islands-nuclear-testing-sea-level-rise/.

Sagan, S. D., & Valentino, B. A. (2017). Revisiting Hiroshima in Iran: What Americans really think about using nuclear weapons and killing noncombatants. *International Security*, *42*(1), 41–79.

Savalei, V., & Rhemtulla, M. (2013). The performance of Robust test statistics with categorical data. *British Journal of Mathematical and Statistical Psychology*, *66*(2), 201–223.

Schuman, H., Ludwig, J., & Krosnick, J. A. (1986). The perceived threat of nuclear war, salience, and open questions. *Public Opinion Quarterly*, *50*(4), 519–536.

Ten Berge, J. M. F., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613–625.

Van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software*, *20*(11), 1–19.

Van de Vijver, F. J. R. & Leung, K. (2011). Equivalence and bias: A review of concepts, models, and data analytic procedures. In D. Matsumoto & F. J. R. Van de Vijver (Eds.), Cross-cultural research methods in psychology. New York: Cambridge University Press.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

Welt Documentary. 2020. The forgotten nuclear war - Bombs on Bikini Atoll [Video]. YouTube. https://www.youtube.com/watch?v=NjqoiT-RS4A.

Wilson, W. (2015). Why are there no big nuke protests? *Bulletin of the Atomic Scientists*, *71*(2), 50–59.

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, *81*(4), 1014–1045.

Zaller, J. R. (1986). *Analysis of information items in the 1985 pilot study*, *Report to the NES Board of Overseers. Center for Political Studies*. Ann Arbor: University of Michigan.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge: Cambridge University Press.

Zweigenhaft, R. L. (1984). What do Americans know about nuclear weapons? *Bulletin of the Atomic Scientists*, *40*(2), 48–50.

Zweigenhaft, R. L., Jennings, P., Rubinstein, S. C., & Van Hoorn, J. (1986). Nuclear knowledge and nuclear anxiety: A cross-cultural investigation. *The Journal of Social Psychology*, *126*(4), 473–484.

Publisher's Note