

VALIDATION OF MEASUREMENT INSTRUMENTS

Open Access



Religiousness worldwide: translation of the Duke University Religion Index into 20 languages and validation across 27 nations

Cecilia Toscanelli^{1,2} , Elizabeth Shino³ , Sarah L. Robinson¹  and Amber Gayle Thalmayer^{1,4*} 

Abstract

Religiousness and spirituality are important in the study of psychology for several reasons: They are central to identity and values; they have been reported as being positively associated with health and well-being; and they capture (and perhaps lead to) the largest measurable psychological differences between societies. At five items, the Duke University Religion Index (DUREL) is an efficient measure, which advantageously distinguishes between religious sentiment and activity, and between formal versus private involvement. This project extends its internal validation throughout the world, with formal tests of measurement invariance in three languages in Namibia (Study 1) and in a global sample of 26 countries (Study 2). Results confirmed a two-subscale factorial structure of Religious Activity (combining organizational and non-organizational activities) and Intrinsic Religiosity in Namibia and in half of the 26-country samples. In 13 other countries, fit was best for a one-factor model. Fit was problematic where there was too little intra-national variance: in China and Japan, where religious involvement is universally low, and in Tanzania, where it is universally high. Scalar measurement invariance was found for the one-factor structure across 13 samples and for the two-factor structure across 11 samples. External validation of the scale is examined using psychological and sociodemographic variables. This validation of the DUREL supports its use across contexts, facilitating increased attention to this important aspect of both personality and culture.

Keywords: Religion, Measurement invariance, Africa, Worldwide, Cross-cultural

Religion is a key aspect of human psychology, which has played an important role in shaping human societies and values (Schulz et al., 2019). Although religious involvement may be decreasing in the USA (Jones, 2021), the majority of people in the world still report being affiliated with a religion (Pew Research Center, 2018), and it is often a central part of their identities (Tarakeshwar et al., 2003). Important to cultural psychology, religiosity and associated values, such as family and gender roles, also demonstrate significant cross-national differences, even

larger in effect than those of popular variables such as individualism and collectivism (Saucier et al., 2015). Furthermore, religious involvement has emerged as a source of resilience against mental disorders (De Berardis et al., 2020), suggesting its potential relevance for understanding national differences in the prevalence of such disorders (e.g., Berkessel et al., 2021; Dückers et al., 2016). Thus, religious sentiment and involvement are important to psychology and cross-cultural research (Tarakeshwar et al., 2003), although these topics have been relatively underrepresented in psychological science, perhaps related to the underrepresentation of researchers from majority world contexts (Thalmayer, Toscanelli, & Arnett, 2021) where religion is more central to daily life (Pew Research Center, 2018).

*Correspondence: ambergayle.thalmayer@uzh.ch; ambergayle@gmail.com

⁴ Department of Psychology, University of Zürich, Binzmühlestrasse, 8050 Zürich, Switzerland
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

There are many ways to measure religiousness (Hall et al., 2008; Remizova et al., 2022). The focus of this study is the popular and highly efficient inventory, the five-item Duke University Religion Index (DUREL), which includes three components: organizational religious involvement, non-organizational involvement (independent prayer, meditation, or study), and intrinsic or subjective religiosity (Koenig & Büssing, 2010). The DUREL has been shown to converge with other commonly used inventories, for example the Santa Clara Strength of Religious Faith Questionnaire (SCSRFQ; Plante et al., 2002), a five-item unidimensional measure emphasizing the force of religious faith (r with DUREL total score .79, Saffari et al., 2013; .86, Storch et al., 2004), and the Personal Religious Inventory (PRI; Lipsmeyer, 1984), with 45 items and subscales for prayer, ritual attendance, and the integration of religion in cognition, affect, and behavior ($r_{ORA, \text{ritual attendance}} = .84$; $r_{NORA, \text{personal prayer}} = .76$; Lace & Handal, 2018).

The DUREL has been used in diverse contexts including Muslim Iran (Saffari et al., 2013), Catholic Portugal (Lucchetti et al., 2012) and in China where multiple religions are practiced (Wang et al., 2014), though without comprehensive validation. In addition to capturing differences between societies in the typical degree of religious involvement (Saucier et al., 2015), the distinctions made by the DUREL between external behavior (practicing religion) and internal sentiment (spiritual feelings) create a potential bridge between cultural and personality psychology, capturing both between-nation and between-person differences (Saucier, 2019). The current project aims to support better inclusion of religious involvement and sentiment in cross-cultural psychological research by introducing 19 total new translations of the DUREL and extending its validation to three languages in Namibia (Study 1), and to 26 countries around the world (Study 2), considering both internal (psychometric properties and cross-cultural measurement invariance) and external (associations with other variables) validation.

Religiousness and culture

Religiousness may be a key variable in the study of cross-cultural differences. In most nations, a majority of people report being affiliated with a religion, but this varies from lows of 13% in China, 28% in the Czech Republic, and 36% in Japan to near 100% in the Middle East, Sub-Saharan Africa, and other parts of Asia (e.g., Malaysia and Indonesia; Pew Research Center, 2018). The five DUREL items had the largest effect sizes for discriminating between 30 national groups among 50 psychological scales hypothesized to distinguish between cultural groups (281 items), including values (e.g., classic contrasts such as individualism and collectivism), world views, behavioral practices, and personality characteristics (Saucier et al., 2015). The

variables that followed the DUREL in terms of effect size included beliefs about family and gender roles, which tend to associate with religious values. In the big picture, using data from the World Values Survey, Schulz et al. (2019) show how the family practice policies of the Western Church may have set in motion lifestyle changes that shaped the psychological changes (individualism, analytic thinking, impersonal pro-sociality) that now define Western culture.

The contemporary significance of inter-cultural differences in religiosity is supported by a study of 106 countries on the importance of religion in peoples' lives (Pew Research Center, 2018). Note that while cross-cultural comparisons must be interpreted with caution, the rating of importance was determined to be the most invariant aspect in a later study, making it the most suitable of those measured for cross-national assessment (Remizova et al., 2022). The world's highest religious importance was reported in Sub-Saharan Africa, where between 71% (Botswana) and 98% (Ethiopia) of respondents reported religion as being "very important" to them. Some countries in Latin America (Honduras and Brazil) had rates nearly as high, while in others half the population or less reported high importance. Middle Eastern countries varied from 36% in Israel to 78% in Iran. For Western countries, the highest rates were in the USA (53%) and Greece (56%), and the lowest in the UK and Germany (10%). Asia-Pacific had the largest within-region contrasts, with high importance varying from 3% in China and 10% in Japan, to 80% in India and 90% in Indonesia. Thus, while religion plays a role in virtually every society, contrasts within and across regions are substantial, defining meaningful differences between groups. It is also noteworthy that the importance of religion tends to be lower in Western, industrialized contexts and higher in the "majority world," e.g., Asia-Pacific, Africa, the Middle East, and Latin America, where almost 90% of humans live, but who are consistently underrepresented in mainstream psychology (e.g., Kagitcibasi, 2002; Thalmayer, Toscanelli, & Arnett, 2021). Better incorporation of religiosity into psychological studies could mean better representing the experiences and perspectives of the global human population.

Since its appearance in the 1990s, the DUREL has been used in many contexts. We identified 16 translations in prior published studies, detailed in Table 1. These include four translations to languages in Asia, three to the Middle East, three to Europe, three to the Americas, and one to Austronesia. None were identified in African languages. Perhaps partly due to the variety of disciplines represented, none of the validation procedures in these studies meet currently accepted standards for the adaptation of psychological measures (e.g., Byrne & van de Vijver,

Table 1 Published translations of the Duke Religion Index by region and language

Region	Language	N	Sample type	Internal structure	Scales used	Convergent validity	Associations	Citation
South America	Portuguese	439	Community	α	Three	-	-	Lucchetti et al., 2012
	French Creole	55	Community	CFA	Total score	Index of Religiousness	Marginality, education, household income	Gonzales et al., 2015
North America	Spanish	30	Community	CFA	Total score	"	"	Gonzales et al., 2015
Asia (South East)	Malay	173	University	EFA, α , test-retest reliability	Total score	-	Religious coping, health, depression, anxiety, stress	Nurasikin et al., 2010
	Thai	800	University	α	Three	-	Eating disorders	Pengpid et al., 2015
Asia (East)	Chinese	2425	Community	α , EFA, CFA, test-retest reliability	Total score	-	-	Chen et al., 2014
	Chinese	1285 & 2564	University & Community	α , PCA, test-retest reliability	Three	-	-	Wang et al., 2014
	Japanese	53	University	CFA	Total score	Index of Relig.	Marginality, education, income	Gonzales et al., 2015
Austronesia	Tagalog	36	Community	CFA	Total score	"	"	Gonzales et al., 2015
Europe	Polish	416	University	α , PCA, test-retest	Total score	-	-	Dobrowolska et al., 2016
	Ukrainian	21	Community	CFA	Total score	Index of Relig.	Marginality, education, income	Gonzales et al., 2015
	Italian	54	Outpatients	None	Total score	-	Suicide ideation	De Berardis et al., 2020
	Portuguese	150	Community	α , EFA	Total score	Belief into Action Scale	-	Martins et al., 2021
	German	123	Psychiatric	α	Three	-	Influence of religiosity/spirituality of psychiatrists on patient's health	Lee & Baumann, 2013
Middle East	Farsi	984	University	α , test-retest reliability	Total score	Hoge Intrinsic Religiosity	-	Hafizi et al., 2014
	Arabic	20	Community	CFA	Total score	Index of Relig.	Marginality, education, income	Gonzales et al., 2015
	Persian	1762 & 796	University	α , EFA, CFA, test-retest reliability	Total score	Strength of Faith Q.	-	Saffari et al., 2013

Note. Emphasis in this table is on translations; the many studies published in English are not included. *PCA* principal component analysis, *CFA* confirmatory factor analysis, *EFA* exploratory factor analysis. We searched for translations in Google Scholar using the name of the index, its acronym, and the terms "validation" or "translation," as well as the names of many commonly used languages in research.

2010; Fischer & Karl, 2019). Six come from a project which tested translations on samples of 20 to 55 individuals, three studies only tested internal consistency, one assessed no psychometric properties, and none tested measurement invariance. Furthermore, while the DUREL was initially presented strictly as a three-component

measure, most later researchers have used it as a unidimensional construct (Chen et al., 2014; Hafizi et al., 2014; Saffari et al., 2013). The internal validation of the DUREL and its proposed structure have thus not been assessed systematically for applicability across contexts. Measurement invariance, in particular, is crucial in cross-cultural

research to ensure that survey scores are appropriately comparable (Fischer & Karl, 2019). This is especially relevant in the study of religiosity where the construct may have highly varying and even culture-specific meanings (Remizova et al., 2022). Establishing the extent of the cross-cultural suitability of this practical inventory and making it available in many languages could allow for better comparisons within and between groups, facilitating exploration of religiousness directly and also making it practical to include as a covariate in studies of other psychological phenomena.

Associations between religiousness and other psychological variables

The association of religious sentiment and/or engagement with other psychological or sociodemographic variables is important to understand, but has so far mainly been studied cross-sectionally, with measures that have not been tested for cross-cultural measurement invariance. The existing literature is briefly reviewed with this caveat in mind, to provide a summary of current assumptions about these associations, and to form loose hypotheses that can be tested in the Namibian and (where possible) global contexts.

Prior research has generally indicated a positive association for religiousness with well-being and health, for example, between religious practice and life satisfaction in a sample of over 20,000 participants (Berthold & Ruch, 2014). Measures of social commitment in religious activities have also indicated a positive association with physical health (Koenig & Larson, 2001; Seybold & Hill, 2001). Meta-analyses associate religiousness with reduced mortality (Chida et al., 2009; McCullough et al., 2000), and better health and longevity (Seybold & Hill, 2001). However, this assumes a “healthy” way of living one’s religion; commitment of a dogmatic or authoritarian type has been linked to intergroup conflict and child abuse (Seybold & Hill, 2001). Associations can also depend on which dimension of religiosity is taken into account: while depressive symptoms are generally lower among the religious (Smith et al., 2003), an “extrinsic” religious attitude (engaging in religious activities for self-serving ends or to avoid dealing with problems) or negative religious coping (blaming God) instead associate with depressive symptoms (Smith et al., 2003).

Studies among Muslims in Farsi (Hafizi et al., 2014) and Catholics in Portuguese (Lucchetti et al., 2012) have reported lower religious engagement among people with more education, using the DUREL total score. A large multi-country study reported that people with more education were less religious in 18 of 39 contexts studied, but that the opposite was true in nine nations, and there was no effect in 12 (Schwadel, 2015). They reported that

association between higher education and lower religiosity was strongest in more religious nations (Schwadel, 2015), but as 26 of the 39 countries were European or closely related Western contexts, the four Asian countries include the wealthiest contexts in the region, and the only African country was South Africa, it is not clear how this finding might generalize to majority world contexts, specifically to Africa, the most consistently religious region in the world. In the only study identified that compared religious involvement to income, scores on DUREL Intrinsic Religiosity were negatively associated with income (Lucchetti et al., 2012).

Women have been assumed to be more religious than men (Argyle & Beit-Hallahmi, 2013), but the religion and the activity may play a role (Vardy et al., 2022). Women identify as more religious in Christian, Muslim, Hindu, and Buddhist, but not Jewish contexts, and engage in more religious activities in Christian, Hindu, and Buddhist, but not Muslim or Jewish contexts (Sullins, 2006). Ultimately, neither gender nor age was found to be a consistent predictor of religiosity in a meta-analysis of 63 studies from 19 countries (Saroglou, 2010).

Goals for the current study

In two studies, we examine the internal and external validation of the DUREL (Koenig & Büssing, 2010) to facilitate its use in the study of psychology within and across cultures. In terms of internal validation, including psychometric and structural validity, in Study 1 we report on the DUREL’s applicability in two African languages, Oshiwambo and Khoekhoegowab, and in English in Namibia. These are the first published translations of the DUREL into African languages, using a multi-step process including expert panels, administered to large samples of community adults. With nearly 98% of the population in Namibia identifying as Christian, and with religious involvement very high throughout Sub-Saharan Africa (Pew Research Center, 2012), this is an important variable to include in local psychological studies. In Study 2, we assess the psychometric and structural properties of the DUREL in the Survey of World Views data, a global sample of university students from 33 countries (described in detail in Saucier et al., 2015; 26 samples used for analysis). As the nations in Study 2 have diverse predominant religions (16 majority Christian, 4 majority Muslim, 1 majority Buddhist, 2 majority Hindu, 2 majority unaffiliated, many with great diversity; detailed percentages are provided in Supplemental Table S1), this allows for the assessment of the instrument in a variety of religious, as well as national and linguistic, contexts. Based on prior research and as reported in our pre-registered analysis plan, we expected the DUREL subscales to have good internal consistency across contexts.

To assess external validation of the DUREL scale, in both studies, for 30 total samples, we examine the association of the DUREL to gender. In the Namibian samples, we additionally examine associations of the DUREL and its subcomponents with age, life satisfaction, physical health, education, income, and employment status. Based on prior literature, we loosely hypothesized finding positive associations between DUREL scores with well-being and physical health, small gender differences in favor of women being more religious, and higher DUREL scores to associate with lower educational level, without a priori expectations regarding the DUREL subcomponents. As prior literature on the relation of religiosity with other aspects of socioeconomic status and age are minimal, we address these associations in an exploratory way.

Together, these studies allow us to assess the internal validity of the DUREL in 30 total contexts, including in 20 total languages, validating 17 new translations, with some additional assessment of external validity, particularly in Namibia. To the extent permissible based on our assessment of cross-sample measurement invariance, we then compare scores across nations. Based on international surveys of religiosity, we expected average scores on the DUREL and/or its subcomponents to be higher in African countries, in India and in Muslim-majority Asian and Middle Eastern countries, and in South and Latin America, and lower in Western contexts to make a cross-cultural assessment of the DUREL's reliability and validity and to draw cross-national contrasts on religious involvement.

Study 1: the DUREL in three languages in Namibia

Method

Participants

Participants were adult native speakers of Oshiwambo in northern Namibia ($n = 678$), native speakers of Khoekhoegowab from villages and towns throughout the country ($n = 645$), and speakers of English (non-native) from the capital city Windhoek and surrounding areas ($n = 589$). Oshiwambo, spoken by nearly 50% of the population, and Khoekhoegowab, spoken by about 12%, are the two most commonly spoken African languages in Namibia (Frydman, 2011). English is the official language since independence in 1990, when it replaced Afrikaans. For this reason, English speakers were recruited in and around the capital where it is more commonly spoken than in rural areas. (Note that while there is a small population of white Namibians of German, English, and Afrikaner heritage, none were included in our samples.) Demographic information collected included age, gender, home language, participant and parents' level of schooling, household income, employment status, and location

of survey-interview. A summary is provided in Table 2; full details are in Supplemental Table S2.

Chi-square tests indicated some significant differences among the samples. Educational attainment, measured by asking to the participants to indicate what is the highest level of education that they completed, was highest among the English-speaking and lowest among the Oshiwambo-speaking sample, $\chi^2(14) = 391.43$, $p < .001$. The same pattern was observed for mother's, $\chi^2(14) = 186.11$, $p < .001$, and father's education, $\chi^2(14) = 173.64$, $p < .001$, and for monthly income, $\chi^2(12) = 317.17$, $p < .001$. To assess employment or engagement in education, participants were asked to choose one or more from seven options. The proportion that reported being a student did not vary by sample, $\chi^2(2) = 2.90$, $p = .23$. However, the English-speaking sample were most likely to have regular part-time, $\chi^2(2) = 38.41$, $p < .001$, or full-time work, $\chi^2(2) = 49.69$, $p < .001$, and the Oshiwambo language sample was least likely. The categories of "currently not working," $\chi^2(2) = 81.34$, $p < .001$, "working at home, or other unpaid work," $\chi^2(2) = 13.26$, $p = .001$, "seeking paid work," $\chi^2(2) = 36.15$, $p < .001$, and "occasional paid work," $\chi^2(2) = 45.26$, $p < .001$, were endorsed most often by Khoekhoegowab speakers and least often by the English language sample.

Procedure

Ethical review of the study plan was made by the University of Namibia's Research Ethics Committee (UREC) and data was collected from July to September, 2019. The English language survey data was collected in a paper and pencil format, as English is the national language and the language of instruction in most Windhoek schools and is both spoken and written in local work settings.

Many potential participants in the Khoekhoegowab and Oshiwambo language samples, however, were expected to lack confidence in reading and writing their mother tongue, because many attended school in Afrikaans or in English despite speaking an African language at home. For this reason, in those samples, the survey was filled out by an interviewer based on the oral responses of participants.

Teams of eight to 15 interviewers/data collectors for each of the three language contexts were graduates of the sociology or psychology programs of the University of Namibia (BA or MA degree), primary- or secondary-school teachers of the language in question, or experienced data collectors, having worked on previous academic studies and/or for the national census survey. Each team met for a weekend-long training led by the second and fourth authors of this report. Data collection occurred in a 4- to 8-week period following the training and was coordinated with local leaders. Participants were

Table 2 Sample characteristics by language group, in percentages

	Khoekhoegowab (n = 645)	Oshiwambo (n = 678)	English (n = 589)	Total (N = 1912)
Percent female	53	58.6	53.3	55.15
Education level completed				
Grade 7 primary or less	14.8	20.1	1.4	12.9
Grade 10 secondary	33.1	27.3	13.7	25.1
Grade 12 secondary	27.7	24.8	26.1	26.2
Some post-secondary or vocational	18.1	18.2	21.9	19.3
University Bachelor's degree	5.4	7.1	27.8	12.9
Master's or post-graduate degree	.9	1.6	9.2	3.7
Employment status				
Not currently working	36.6	25.7	13.9	25.7
Students	21.6	21.2	24.6	22.4
Work at home, other unpaid work	10.4	6.8	5.3	7.5
Seeking paid work	18.6	12.7	7.3	13
Occasional paid work	14.9	6.6	5.3	8.9
Regular part-time paid work	7.9	1.5	9	6
Regular full-time paid work	35	28.9	47.5	36.7
Median monthly income, in US \$				
Under US\$35	41	56.3	25.6	41.6
Between US\$35 and US\$200	29.8	24.6	19.5	24.8
Between US\$200 and US\$675	20.1	11.6	23.1	17.9
Over US\$675	9.2	7.6	31.8	15.6

Note. Age for the Khoekhoegowab-speaking sample $M = 34.8$, $SD = 11.1$; Oshiwambo-speaking $M = 33.3$, $SD = 12.1$; English-speaking $M = 31.7$, $SD = 9.6$

recruited by interviewers in their home communities from among neighbors, church members, colleagues, the parents of students, and strangers from nearby villages and neighborhoods. Interviewers asked participants their age and their home language, and they noted gender, the location of the survey, and notes about how the participant was recruited and how the interview went. The interview and the preliminaries regarding informed consent and instructions were in the same language as the survey. Written informed consent was obtained from all participants. The interview typically lasted approximately 40 min.

Materials

Results from three inventories are reported here. The full survey also included inventories on mental health and personality traits, described in other projects and in the pre-registered analysis plan, https://osf.io/y8d4z/?view_only=347d245ea9544b09a4111a8650059af4. All study materials are available at https://osf.io/6d8gs/?view_only=a32710f4784e447f890aa077ff6b5bc9.

The inventories were translated into Khoekhoegowab and into Oshiwambo (Oshikwanyama dialect) following a process defined by the World Health Organization. This is described in detail here, as potential guide

for other researchers. While such a process is probably always ideal, it is arguably especially relevant when working in smaller-scale languages, for example as there is no published English to Oshiwambo dictionary. (1) *Forward translation* was completed for Khoekhoegowab by a native speaker and PhD candidate in African linguistics, and for Oshiwambo by a native speaker and professional translator referred to us by the African Languages department at the University of Namibia. (2) *Expert panels* were held for each language, including the initial translators, subject matter experts and language experts all of whom were mother-tongue speakers, and the last author of this report who was the principal investigator of the study. For Khoekhoegowab, the additional experts included a clinical psychologist, a social worker, and a senior lecturer of the Khoekhoegowab at the University of Namibia. For Oshiwambo, in addition to the original translator, they included four Oshiwambo-speaking faculty members of the University of Namibia: a clinical psychologist and senior lecturer of psychology (the second author of this report), a lecturer of psychology, a lecturer of education, and a senior lecturer of Oshiwambo. At these day-long panel meetings, the initial forward translation next to the original English was projected onto a screen. The group discussed every line of translation, ultimately coming to

consensus on a refined translation. (3) *Back-translation* of the refined translation was conducted by a professional translator, in each case, who had not been present at the panel meeting and had no prior knowledge of the surveys. This version was reviewed together with the expert, the initial translator, and the third and final authors of this report, and adjustments were made. (4) *Pre-testing and cognitive interviewing* were conducted by a research assistant in each of the languages, who piloted the survey with members of the relevant communities. These translations are made freely available to other researchers in the online materials posted for this study.

Note that because the English language version of the survey was completed on paper, items were worded in the first person. The Khoekhoegowab and Oshiwambo surveys were planned to be read aloud to participants, so items were worded in the second person.

Duke University Religion Index (DUREL) The five items measure three aspects of religiosity: Organizational Religious Activity (ORA; one item), Non-Organizational Religious Activity (NORA; one item), and Intrinsic Religiosity (IR; three items; Koenig & Büssing, 2010). Responses on a 6-point Likert-type scale are linked to different terms depending on item (respectively: “never” to “more than once a week;” “rarely or never” to “more than once a day;” “definitely not true of me” to “definitely true of me”). Though the authors recommend against a total score, EFA and CFA support for one has been reported in samples from Iran (Saffari et al., 2013) and China (Chen et al., 2014), and many other studies have used one. The mean, standard deviation, Cronbach’s alpha, and Omega

of the DUREL in the three languages are reported in Table 3. To provide comparability with prior research and facilitate reproducibility for future researchers, and with the justification of high alpha and Bollen’s omega coefficients, we report all raw item, subscale, and total scores.

Satisfaction with Life Scale (SWLS) This 5-item measure assesses well-being in terms of global cognitive judgment (Diener & Emmons, 1985). The scale showed acceptable reliability in all languages (English $\alpha = .74$; Khoekhoegowab $\alpha = .74$; Oshiwambo $\alpha = .70$). The measurement invariance of the SWLS across these samples is described in detail in a separate report (CITATION REDACTED FOR BLIND REVIEW). Support was found for a 4-item version, excluding the fifth item, which is more abstract than the others and has been found to be problematic in many studies. The analyses here rely on the first four items only.

General Self-Reported Health (GSRH) Physical health was rated on a 5-point scale from poor to excellent. A meta-analysis of 22 studies has shown this one-item self-assessment to correlate highly with longer and more invasive measures of health status and to be strongly associated with risk of death over 5 years (DeSalvo et al., 2006; DeSalvo & Muntner, 2011).

Analyses

Data exclusions Based on criteria described in the pre-registered analysis plan, 125 of the 2037 cases collected

Table 3 Descriptive statistics and psychometric properties for DUREL items and components

Item or scale	English <i>n</i> = 573		Khoekhoe <i>n</i> = 645		Oshiwambo <i>n</i> = 672	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
How often do you attend church, or other religious services? (ORA) ^a	3.88	1.37	3.73	1.52	3.66	1.23
How often do you spend time in private religious activities, such as prayer, meditation, or study of religious scriptures? (NORA) ^b	3.40	1.75	3.83	1.66	2.58	1.30
In my life I experience the presence of the Divine. (IR1) ^c	4.00	1.13	4.29	1.22	3.71	1.43
My religious beliefs are what really lie behind my whole approach to life. (IR2) ^c	3.86	1.28	4.17	1.09	3.96	1.28
I try hard to carry my religion over into all other dealings in life. (IR3) ^c	3.70	1.31	4.14	1.12	3.84	1.39
Religious Activity (ORA & NORA)	3.62	1.38	3.78	1.37	3.13	1.15
Intrinsic Religiosity (IR)	3.85	.97	4.20	.95	3.84	1.22
Total score	3.76	.96	4.03	.95	3.56	1.00
α Total score	.78		.82		.85	
ω Total score	.78		.82		.85	

Note. ω = Ordinal omega (Bollen). * $p < .05$; ** $p < .01$; *** $p < .001$

^a Possible answers range from 0 (never) to 5 (more than once a week)

^b Answers range from 0 (rarely or never) to 5 (more than once a day)

^c Answers range from 0 (definitely not true of me) to 5 (definitely true of me)

were excluded from the analyses. This included those marked for exclusion by the interviewer (either because it was not completed or because the interviewer did not believe it was reliably completed; $n = 15$); from participants under age 18 ($n = 2$); missing more than 15% of total item responses ($n = 21$); or where problems were found in assessment of the longer questionnaires, for example where there was almost no variation in response option usage across 30 or more items ($n = 87$).

Missing data The MissMech package in R (Jamshidian et al., 2014) was used to evaluate the homoscedasticity, multivariate normality, and missing completely at random (MCAR) status of the data in preparation for analysis (Jamshidian & Jalal, 2010). While multivariate normality was rejected, there was no reason a priori to expect normality given the skewed ordinal distribution of the DUREL. The nonparametric test of homoscedasticity was not rejected ($p = .24$), and there was not sufficient evidence to reject MCAR at $p < .05$. The fraction of missing information (fmi), e.g., how much estimation is affected by nonresponses (Savalei & Rhemtulla, 2012), for item means and variances was less than .01, indicating little to no impact of missing data on obtained estimates. Together, these analyses gave us confidence in our assessment that missing data were likely missing at random, if not missing completely at random, and the impact of missing values would likely be negligible. Multiple imputation or estimates relying on FIML were therefore utilized. Multivariate nonnormality, however, led us adopt distribution-free methods for CFA and SEM modeling, specifically WLSMV (Blunch, 2012).

Psychometric properties and structure We tested the factor structure of the DUREL for each language group using exploratory factor analysis (EFA) to confirmatory factor analysis (CFA) with a random split training/testing design. Parallel analysis (O'Connor, 2000), Velicer's minimum average partial (MAP) test (Velicer, 1976), VSS complexity 1 and 2, empirical BIC, and sample size-adjusted BIC were used to assess the optimal number of factors. For CFA, the criteria used to compare the models followed standard practice by considering multiple indices of acceptable fit (Bowen & Masa, 2015; McDonald & Ho, 2002; Vandenberg & Lance, 2000), including a decrease in chi-square value; Tucker-Lewis index (TLI; Tucker & Lewis, 1973) and comparative fit index (CFI) values over .95 (Cheung & Rensvold, 2002; Hu & Bentler, 1999); McDonald's noncentrality index (MFI; McDonald, 1989; McDonald & Marsh, 1990) over .90; and root mean square error of approximation (RMSEA; Steiger, 1990) and standard root mean square residual (SRMR; Bentler, 1995) below .08 (Cheung & Rensvold, 2002; Hu & Bentler, 1999). Because chi-square and RMSEA can be

biased by degrees of freedom or sample size (Chen et al., 2008; Shi et al., 2020), we interpret them in light of other indices. It is recommended that when degrees of freedom are low, as here, higher RMSEA should not be taken as indicative of problematic fit in the absence of problems with other fit indices (Kenny et al., 2015).

Measurement invariance and group differences in religiosity We tested the cross-group invariance of the DUREL using multi-group confirmatory factor analysis (MG CFA) with lavaan (Rosseel, 2012) and associated packages in R (R Core Team, 2020). Following standard procedures with recommended modifications for ordinal data (Bowen & Masa, 2015; Byrne & van de Vijver, 2010; Fischer & Karl, 2019; Putnick & Bornstein, 2016), we assessed for three levels of invariance: (1) configural, with parameters free to vary for each group; (2) metric, constraining factor loadings to be equal; and (3) scalar, constraining both item loadings and thresholds to equality across groups. The location and scale of each latent item response underlying the five ordinal indicators were identified using delta parameterization and identification constraints recommended by Wu and Estabrook (2016). We used polychoric correlation and weighted least squares estimation (WLVS) with robust standard errors and a mean and variance adjusted test statistic (Satterthwaite approach; Rosseel, 2012). Again, multiple indices were used to judge model comparison (Bowen & Masa, 2015; McDonald & Ho, 2002; Vandenberg & Lance, 2000), and the chi-square test, which is known to be sensitive to sample size and can lead to over-rejecting invariance, is interpreted in light of TLI, CFI, MFI, RMSEA, and SRMR. The difference in fit across levels of invariance was assessed based on a difference in CFI $< .01$ and a change in RMSEA $\leq .01$ (Svetina et al., 2020). Group latent means can be compared using ANOVA with LSD post hoc tests where group equivalence at the scalar level is established.

Associations with other variables For each language group, we were interested in how sociodemographic variables predict religiosity, and in how religiosity predicts physical and mental health. We thus report standardized coefficients from a regression of DUREL subcomponents on sociodemographic variables (age, gender, education, and income on), and of SWLS and GSRH scores on DUREL subcomponents, accounting for the sociodemographic variables. For these analyses, education was recoded into a continuous variable indicating years of study: did not finish primary school = 0, grade 7 = 7, grade 10 = 10, grade 12 = 12, vocational or other diplomas = 14, bachelor's degree = 16, master or postgraduate = 18. Similarly, the categories used to assess

income were recoded into a continuous variable using the mid-point of the ranges in Namibian dollars that were provided (see Supplemental Table S3; the new coding includes values of 0, 250, 1000, 2250, 4000, 7500, and 10,000).

To evaluate associations with the categorical variable of employment, we employed ANOVA with Tukey’s post hoc tests. From among several options to describe employment, four mutually exclusive categories were created: student; working part time; working full time; unemployed (including those seeking work or working only occasionally). Those who selected only “work at home or other unpaid work” were excluded from this analysis. Due to making multiple comparisons, we avoid the standard $p < .05$ criteria and instead only interpret ANOVA differences and Tukey’s HSD post hoc tests where $p < .01$.

Results

Exploratory to confirmatory factor analyses

For all three samples, Bartlett’s sphericity measures and Kaiser, Meyer, Olkin (KMO) measures of sampling adequacy confirmed the data to be appropriate for factor analysis: English $\chi^2(10) = 637.54, p < .001, KMO = 0.72$; Khoekhoegowab $\chi^2(10) = 861.56, p < .001, KMO = 0.76$; Oshiwambo $\chi^2(10) = 1640.22, p < .001, KMO = 0.70$. Parallel analysis, VSS complexity 1 and 2, empirical BIC, sample size-adjusted BIC, and Velicer’s MAP test indicated a two-factor configuration for all three samples, with results reported in Supplemental Table S2. We considered retaining a model of three components based on the theoretical foundation of the DUREL and face validity (Wang et al., 2014), but CFA with two single-item factors was undermined by lack of identifiability, and EFA results suggested that the religious activity items (ORA and NORA) optimally loaded together as one factor, leading us to rely on a two-factor model for CFA. CFA confirmed this, with improvement in all fit indexes, as

shown in Table 4 (the same analyses using scaled values are reported in Supplemental Table S5). The two-factor configural models with item loadings and factor correlations are displayed graphically in Fig. 1.

Measurement invariance

Results of measurement invariance analysis among the three language groups are reported in Table 5. With increasing constraints, goodness-of-fit indices remained stable, with RMSEAs at .06, CFI and TFI > .99, MFI > .97, and SRMR < .04. Decrease in CFI and TFI and other fit indices were negligible with increasing parameter constraints (Cheung & Rensvold, 2009). From these results, we conclude that invariance was supported at the scalar level: factor location and scale as well as item loadings and thresholds are equivalent between groups, allowing direct comparison of scores. Item thresholds established for Namibia across language groups for each item, grouped by scale, are shown in Fig. 2.

Group differences in religiosity

Because scalar level invariance was established for the two-component model, we compared mean scores across groups using latent scores using ANOVA, with results reported in Supplemental Table S4. No significant differences in either religious activity or intrinsic religiosity were seen between the three language groups.

Associations with other variables

In Table 6, we report the standardized coefficients from a regression of latent scores on the DUREL sub-components on sociodemographic variables (age, gender, education, and income on), and of SWLS and GSRH scores on DUREL sub-components (latent scores), accounting for the sociodemographic variables. The most consistent predictor of religious engagement was gender: Women scored higher in all three groups, on both subscales for English and Khoekhoegowab language groups, and for religious activity for the Oshiwambo group. Older participants in the

Table 4 DUREL configural fits: one vs two factors in three languages in Namibia

Language	Factors	χ^2	df	p	RMSEA	CFI	TLI	SRMR	MFI	ΔCFI	ΔTLI
English	1	94.26	5		.177	.950	.900	.087	.925		
	2	12.75	4	<.001	.062	.995	.988	.034	.992	.045	.088
Khoekhoe	1	63.56	5		.135	.977	.953	.072	.956		
	2	4.16	4	<.001	.008	1.00	1.00	.020	1.00	.023	.047
Oshiwambo	1	345.24	5		.318	.968	.936	.144	.776		
	2	21.76	4	<.001	.081	.998	.996	.042	.987	.030	.060

Note. The two-factor model includes one factor (Religious Activity) composed of two items (ORA and NORA). The second factor (Intrinsic Religiosity) includes three items

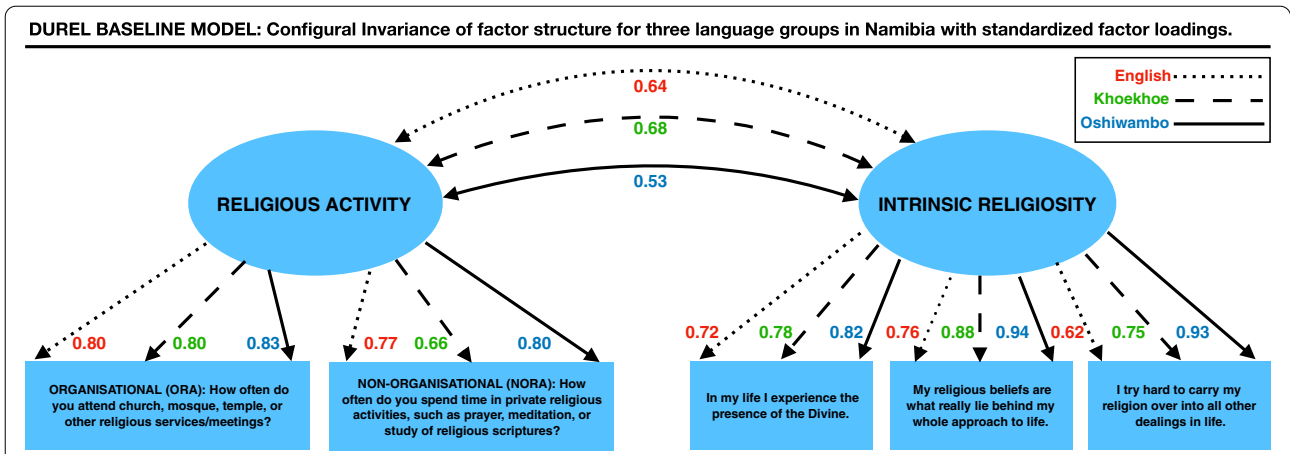


Fig. 1 CFA loadings for a two-factor model of the DUREL in three languages

Table 5 Model indices for measurement invariance of the DUREL in three languages in Namibia

	χ^2	df	RMSEA	RMSEA 90% CI	CFI	TLI	SRMR	Δ FI	Δ CFI	Δ TLI
Configural	38.67	12	0.059	[0.039,0.081]	.998	.996	.032	.993	-	-
Metric	62.34	18	0.063	[0.046,0.080]	.997	.995	.040	.988	-.001	-.001
Scalar	144.37	42	0.062	[0.051,0.074]	.993	.995	.038	.973	-.004	.000

Note. For configural invariance, parameters are free to vary for each group; for metric, factor loadings are constrained to be equal; and for scalar, both item loadings and thresholds are constrained to equality across groups

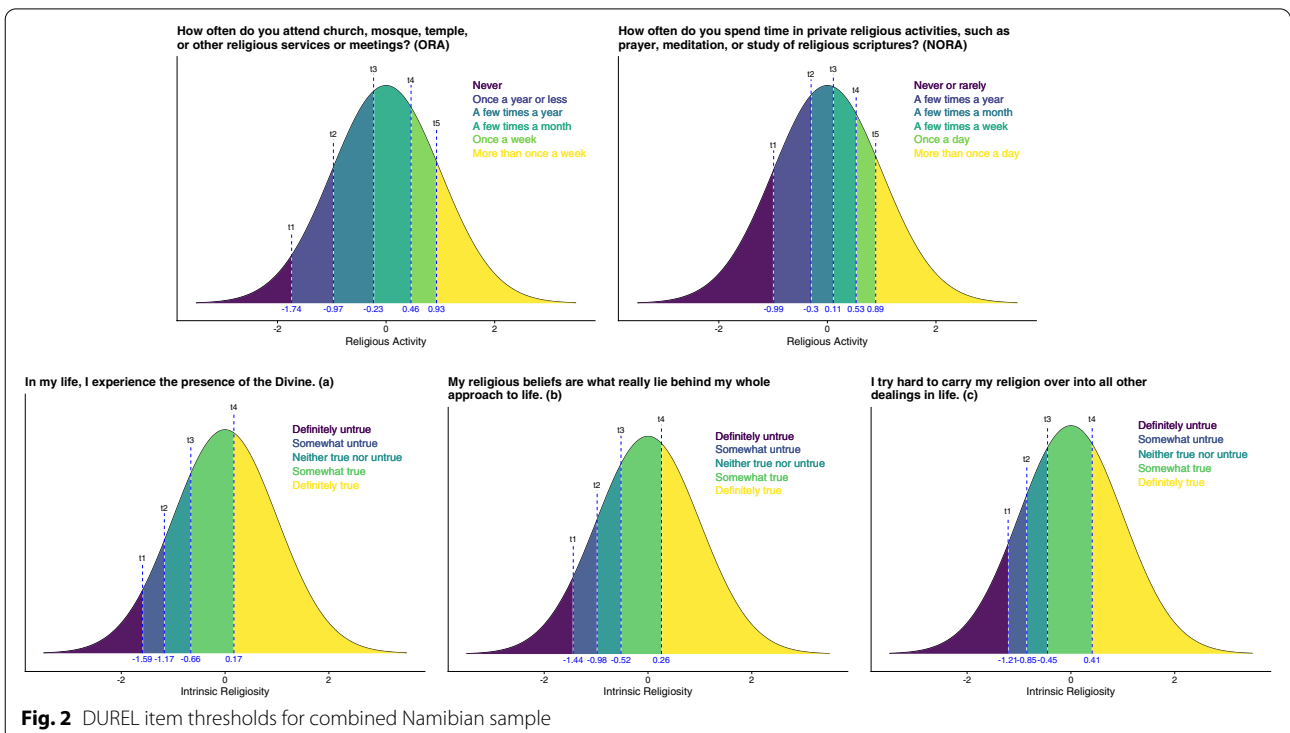


Fig. 2 DUREL item thresholds for combined Namibian sample

Table 6 Standardized coefficients for associations of latent scores on DUREL scales with life satisfaction, health, and demographic variables

	English		Khoekhoegowab		Oshiwambo	
Sociodemographic Variables as Predictors of DUREL Subscales						
	RA	IR	RA	IR	RA	IR
Age	ns	ns	.12*	.12*	ns	.15**
Gender (Male)	-.20**	-.20**	-.15**	-.16**	-.12*	ns
Education	.15*	.17**	ns	ns	ns	ns
Income	ns	ns	ns	ns	ns	ns
DUREL Subscales as Predictors of Health and Satisfaction						
	SWLS	GSRH	SWLS	GSRH	SWLS	GSRH
Age	ns	-.18**	ns	-.13*	ns	-.20**
Gender	ns	ns	ns	.11*	ns	ns
Education	ns	ns	.12*	.18**	.13*	.16**
Income	.24**	.23**	.27**	ns	.18**	.12*
RA	ns	ns	.19*	ns	ns	.15*
IR	ns	ns	ns	.21**	.28**	-.13*

Note. RA Religious Activity, IR Intrinsic Religiosity, SWLS Satisfaction with Life Scale, GSRH General Self-Reported Health

* $p < .01$; ** $p < .001$

Khoekhoegowab and Oshiwambo samples were more religious, but this was not true in the English language sample. Higher education was associated with higher religiosity for the English language group, but not for the other groups. There were no significant associations with income.

Satisfaction with Life was predicted by Religious Activity for Khoekhoegowab speakers only, and by Intrinsic Religiosity for Oshiwambo speakers only. Self-reported health was predicted by Religious Activity for Oshiwambo speakers only. Intrinsic Religiosity predicted better self-reported health for Khoekhoegowab speakers but worse health for Oshiwambo speakers. Results were more consistent for the sociodemographic variables in the model, with higher age predicting worse health in all three samples, and higher education and income generally associating with better satisfaction and health.

Differences in religiosity based on level of employment (student, working part time, working full time, or unemployed) were tested using ANOVA, interpreting differences with post hoc tests only where significant at $p < .01$. For both English and Oshiwambo speakers, there were no significant differences at this level. For Khoekhoegowab speakers, there were significant but small effects indicating more religious involvement among those who are employed full-time versus those who are unemployed, for RA $F(1, 4) = 4.19, p < .01, \eta^2 = .026$, and IR, $F(1, 4) = 5.57, p < .05, \eta^2 = .034$.

Discussion

In Study 1, we tested the DUREL's psychometric properties, structure, and association with other variables for three language groups in Namibia, including the two most-spoken African languages and the national language of English. These two new translations, the first into African languages, were developed using a multi-step translation process including expert panels and were tested on large samples of community adults representative of the local populations. They are now freely available to the scientific community. Given that Africa may be the most religious region in the world (e.g., Pew Research Center, 2018), this study of the DUREL in an African context, to our knowledge the first and only, is overdue.

Consistent with prior research but contrary to the recommendations of the original authors, we found the DUREL total score to have good internal consistency. However, our analysis with multi-group CFA better supported a two-factor structure. The three-item factor of Intrinsic Religiosity was well supported, while the single-item scales for Organizational and Non-Organizational religious activity loaded together as one factor, which we termed "Religious Activity." In Namibia, at least, it appears that those who attend church also tend to spend time in private religious activities. We were able to establish scalar level measurement invariance for this two-factor model across the three groups, indicating that the scale is used similarly, and scores can be compared, across these groups in Namibia. We did not find

differences in Religious Activity or Intrinsic Religiosity among the groups.

Association of the DUREL with life satisfaction, health, and sociodemographic variables partially confirmed hypotheses based on prior literature. As has been seen in other contexts, more religious Namibians reported higher life satisfaction and health in some regards, though these results were not consistent across DUREL subcomponents or samples, with more non-significant than significant associations. Contrary to prior findings (e.g., Hafizi et al., 2014; Lucchetti et al., 2012; Schwadel, 2015), more educated people were more rather than less religious in the English language Namibian sample, though there was no significant association in the other samples, after controlling for other sociodemographic variables. Given the generally high religious engagement in Namibia, this finding in particular contradicts the finding of Schwadel (2015) that the role of education in increasing secularization is especially strong in highly religious contexts. This would need to be replicated in other samples before conclusions can be drawn, but suggests that the role of education may differ in under-studied, majority world contexts. There were no significant associations with income.

As has been seen in other Christian samples, women reported more religious involvement than men, both in terms of religious activity and private sentiment. Exploratory analyses found that older participants were more religious in some regards. Being employed full-time, a variable not hitherto tested in relation to the DUREL, was associated with higher religious involvement only among Khoekhoe speakers.

Study 2: the DUREL in 26 nations

Method

Participants and materials

The following analyses used the Survey of World Views data, the full sample of which includes 8887 college-student participants from 33 countries in nine regions of the world. Full demographic characteristics and the recruitment strategy, as well as the full list of variables administered, are described by Saucier et al. (2015). The nations included in the current study are those for which there were a minimum of 100 cases with complete responses on the DUREL. These 26 samples are detailed in Table 7 with means and standard deviations for the total score and subscales, and alpha and omega values for the total score, to facilitate comparison with other published results.

Analyses

To assess the internal validity of the DUREL in terms of psychometric properties and structure, and its

measurement invariance across nations, we followed the procedures and criteria detailed for Study 1. We anticipated establishing configural and full- or partial-metric invariance, indicating similar form and contribution of items to the construct across groups. However, because the number and diversity of groups was large, we expected we might find the DUREL to be ill-suited to some samples, and planned to exclude samples from group analysis where baseline configural-fit in CFA was problematic (Byrne & van de Vijver, 2010). Where scalar invariance was established (i.e., equivalence of both item loadings and thresholds) for either a one- or two-factor structure, comparison of latent means was made using ANOVA (Kim et al., 2017; Putnick & Bornstein, 2016). Finally, we evaluated the association between religious involvement and gender using ANOVA. Education and age were not tested here because all participants were college students, making for minimal variation in the sample.

Results

Structure

Nation-specific CFA fit indices for one- and two-factor models are reported in Table 8. Based on outcomes, two groups were designated. In 10 countries (Bangladesh, Ethiopia, Greece, India, Kenya, Morocco, Nepal, Philippines, Taiwan, Thailand), the two-factor model had better fit and correlations between the factors below .85, indicating a difference in meaning between the two constructs (Kenny, 2016). For 11 countries (Argentina, Brazil, Canada, UK, Germany, Peru, Poland, Singapore, Spain, Ukraine, USA), both models had adequate fit; in most cases, goodness-of-fit indices for the two-factor configuration outperformed one-factor indices. For these countries, correlations between the two factors were greater than .85, indicating substantial shared variance (Kenny, 2016). A preference for the most parsimonious solution led us to select a one-factor model for this group. For two samples, Turkey and Malaysia, results for the two-factor model were unreliable due to high covariance between latent factors, and we classed these nations in the one-factor group. In this group, the difference between Intrinsic Religiosity and Religious Activity was not large enough to justify two factors. Three samples (China, Japan, and Tanzania) were excluded from either group. A lack of variation within these samples, due to extremely low religiosity in China and Japan and to consistently high scores in Tanzania, led variance-covariance matrices of estimated parameters to be non-positive definite.

Measurement invariance

Results of the two sets of measurement invariance analyses are reported in Table 9. The one-factor structure was

Table 7 Sample details, descriptive statistics, and psychometric properties for DUREL items and subscales in 26 countries

Region, nation	Language	n	Female%	M _{age}	ORA		NORA		RA		IR		Total score			
					M	SD	M	SD	M	SD	M	SD	M	SD	α	ω
Africa (Sub-Saharan)																
Ethiopia	Amharic	381	29	24.0	4.57	1.46	3.85	1.51	4.21	1.36	4.09	1.02	4.14	1.05	.90	.90
Kenya	English	288	33	24.6	4.44	1.19	3.75	1.44	4.08	1.21	3.69	1.00	3.85	0.96	.87	.87
Tanzania	Kiswahili	256	33	24.8	5.10	0.76	4.77	1.09	4.94	0.83	3.82	0.95	4.27	0.79	.77	.76
North Africa/Middle East																
Morocco	Arabic	441	50	25.6	4.68	1.51	5.16	1.45	4.91	1.24	4.38	0.83	4.59	0.84	.77	.77
Turkey	Turkish	416	54	21.1	2.42	1.52	3.09	1.84	2.75	1.50	2.94	1.31	2.86	1.32	.94	.94
South Asia																
Bangladesh	Bengali	272	22	21.7	4.36	1.55	3.67	1.49	4.02	1.32	3.53	1.07	3.73	0.98	.81	.80
India	English	390	62	21.1	3.54	1.32	3.25	1.60	3.39	1.28	3.12	1.04	3.22	1.01	.81	.81
Nepal	Nepali	346	59	21.0	3.85	1.16	3.44	1.38	3.65	1.14	3.42	0.96	3.51	0.89	.81	.81
Southeast Asia																
Malaysia	Malay	324	66	20.5	4.22	1.32	5.22	1.41	4.71	1.12	4.69	0.67	4.70	0.78	.92	.93
Philippines	Tagalog	425	68	20.0	4.55	1.08	4.37	1.39	4.46	1.08	4.00	0.96	4.18	0.89	.86	.86
Thailand	Thai	350	72	21.6	3.65	1.21	3.59	1.35	3.62	1.09	3.26	0.88	3.40	0.81	.74	.75
Singapore	English	304	55	21.7	2.93	1.56	2.37	1.65	2.65	1.50	2.87	1.27	2.78	1.28	.94	.94
East Asia																
China	Chinese	350	73	20.8	1.76	0.82	1.26	0.63	1.51	0.63	1.93	0.88	1.76	0.69	.88	.88
Japan	Japanese	429	63	20.9	1.44	0.80	1.24	0.69	1.38	0.65	1.67	0.71	1.54	0.54	.84	.85
Taiwan	Chinese	395	64	22.6	2.58	1.08	1.57	0.97	2.07	.93	2.54	1.05	2.35	0.90	.90	.90
Eastern Europe																
Greece	Greek	246	70	21.8	2.70	0.91	2.35	1.53	2.53	1.10	2.34	1.14	2.44	1.03	.91	.91
Poland	Polish	225	88	21.2	3.11	1.39	2.31	1.59	2.70	1.38	2.69	1.33	2.70	1.27	.94	.94
Ukraine	Ukrainian	244	64	20.2	2.65	1.08	2.43	1.57	2.54	1.19	2.88	1.15	2.74	1.05	.89	.89
Western Europe																
Germany	German	349	52	23.6	2.23	1.01	1.65	1.23	1.94	1.00	1.94	1.13	1.94	1.00	.94	.94
Spain	Spanish	379	64	22.7	1.76	1.08	1.54	1.20	1.63	1.02	1.68	0.99	1.66	0.94	.94	.94
UK	English	229	62	22.7	2.09	1.33	1.96	1.54	2.02	1.35	1.98	1.27	2.00	1.26	.97	.97
North America																
Canada	English	220	61	21.8	2.57	1.58	2.28	1.65	2.43	1.54	2.51	1.37	2.48	1.36	.96	.95
USA	English	425	57	21.9	2.99	1.55	2.78	1.72	2.87	1.50	2.87	1.40	2.87	1.37	.95	.95
South America																
Argentina	Spanish	243	56	24.3	2.45	1.38	2.28	1.65	2.38	1.33	2.70	1.12	2.57	1.09	.87	.88
Brazil	Portuguese	195	79	22.2	3.25	1.61	3.00	1.68	3.12	1.49	3.31	1.27	3.24	1.29	.93	.93
Peru	Spanish	309	61	21.8	2.76	1.33	2.47	1.65	2.62	1.31	2.97	1.15	2.83	1.11	.89	.90

Note. ORA Organizational Religious activity, NORA Non-Organizational Religious Activity, RA Religious Activity (ORA and NORA), IR Intrinsic Religiosity

assessed across 13 nation-samples as described above. The two-factor structure was assessed among 13 samples, including the three groups from Namibia, in addition to the 10 nations noted above. For both analyses, invariance was supported at the scalar level, with loadings and thresholds constrained between groups. With increasing invariance constraints, goodness-of-fit indices remained rather stable. Drops in CFI and TFI and other alternative fit indices with increasing parameter constraints were negligible (Cheung & Rensvold, 2009; Putnick &

Bornstein, 2016). From these results, we conclude that factor location and scale as well as item thresholds are equivalent between groups, allowing direct group comparisons of latent means and variances in religiosity and their subscales among the samples within each set.

Group differences in religiosity

In Fig. 3, DUREL latent scores are arranged from highest to lowest for the two sets countries with the three samples from Namibia from Study 1 combined into a single

Table 8 CFA fit indices for one- and two-factor models in 26 countries, grouped by the best fitting model

Nation	Model	r_{RA-IR}	χ^2	<i>df</i>	<i>p</i>	RMSEA	CFI	TLI	SRMR	MFI
Two-factor model group										
Bangladesh	1	-	78.43	5	<.001	.245	.946	.891	.107	.861
	2	.56	4.77	4	.312	.028	.999	.999	.030	.998
Ethiopia	1		63.22	5	<.001	.182	.989	.979	.063	.921
	2	.78	2.39	4	.665	.000	1.00	1.00	.015	1.00
Greece	1		20.48	5	.001	.116	.995	.990	.050	.967
	2	.83	4.88	4	.300	.031	1.00	.999	.023	.998
India	1		26.29	5	<.001	.114	.984	.967	.056	.968
	2	.80	10.78	4	.029	.072	.995	.987	.038	.990
Kenya	1		111.71	5	<.001	.282	.968	.935	.107	.820
	2	.71	29.37	4	<.001	.154	.992	.981	.054	.954
Morocco	1		23.94	5	<.001	.097	.981	.961	.066	.977
	2	.69	6.20	4	.185	.037	.998	.994	.039	.997
Namibia	1	-	781.19	47	<.001	.158	.951	.969	.110	.823
	2	.41	144.36	42	<.001	.062	.993	.995	.038	.973
Nepal	1		56.11	5	<.001	.189	.961	.922	.090	.915
	2	.66	7.03	4	.134	.052	.998	.994	.036	.995
Philippines	1		39.75	5	<.001	.144	.990	.980	.068	.950
	2	.72	4.22	4	.377	.013	1.00	1.00	.023	1.00
Taiwan	1		54.60	5	<.001	.173	.993	.985	.068	.928
	2	.75	2.68	4	.614	.000	1.00	1.00	.016	1.00
Thailand	1		46.13	5	<.001	.161	.966	.932	.070	.937
	2	.63	1.39	4	.845	.000	1.00	1.01	.015	1.00
One-factor model group										
Argentina	1	-	3.97	5	.553	.000	1.00	1.00	.030	1.00
	2	.92	2.35	4	.672	.000	1.00	1.00	.023	1.01
Brazil	1	-	13.11	5	.022	.093	.998	.996	.037	.979
	2	.96	12.10	4	.017	.103	.998	.995	.035	.979
Canada	1	-	30.50	5	.000	.154	.998	.996	.043	.942
	2	.89	1.78	4	.777	.000	1.00	1.00	.012	1.01
Germany	1		11.41	5	.044	.063	.999	.998	.030	.990
	2	.90	.65	4	.957	.000	1.00	1.00	.007	1.01
Malaysia ^a	1		6.13	5	.294	.028	1.00	1.00	.027	-
Peru	1		11.54	5	.042	.079	.998	.995	.039	.985
	2	.90	6.02	4	.198	.049	.999	.998	.027	.995
Poland	1		37.71	5	<.001	.172	.996	.991	.049	.929
	2	.91	26.82	4	<.001	.160	.997	.992	.039	.950
Singapore	1		16.69	5	.005	.095	.999	.997	.030	.978
	2	.91	2.21	4	.698	.000	1.00	1.00	.012	1.00
Spain	1		9.41	5	.094	.053	.999	.998	.029	.993
	2	.93	4.91	4	.297	.027	1.00	1.00	.021	.999
Turkey ^a	1		16.20	5	.006	.077	.999	.998	.027	
Ukraine	1		17.93	5	.003	.116	.993	.986	.053	.967
	2	.85	9.61	4	.048	.085	.997	.992	.036	.986
UK	1	-	13.02	5	.023	.085	1.00	.999	.021	.982
	2	.97	11.30	4	.023	.091	1.00	.999	.018	.984
USA	1		4.95	5	.422	.000	1.00	1.00	.014	1.00
	2	.96	2.07	4	.724	.000	1.00	1.00	.009	1.00

Note. RA Religious Activity (ORA and NORA)

^a Lack of reportable fit for two-factor configuration due to singularity of the two latent factors

Table 9 Measurement invariance of the DUREL for one-factor and two-factor groups

	χ^2	df	RMSEA	[90% CI]	CFI	TLI	SRMR	MFI	Δ CFI	Δ TLI
One-factor structure group, including 13 nation-samples										
Configural invariance	192.571	65	0.087	[.073, .101]	.999	.997	.031	.981	-	-
Metric Invariance	420.629	113	0.102	[.092, .113]	.997	.996	.051	-.955	-.002	-.001
Scalar Invariance	578.622	257	0.069	[.062, .077]	.997	.998	.037	.953	0	.002
Two-factor structure, including 11 countries (13 groups)										
Configural invariance	112.376	52	0.055	[.041, .069]	.999	.996	.030	.994	-	-
Metric Invariance	385.251	88	0.094	[.084, .104]	.993	.990	.056	.971	-.006	-.006
Metric Invariance	748.559	232	0.076	[.070, .082]	.988	.993	.045	.949	.005	.003

Note. Metric invariance means constraining loadings to be equal; scalar refers to constraining both loadings and thresholds to be equal. Based on CFA results, the one-factor structure was tested for invariance across 13 samples: Argentina, Brazil, Canada, UK, Germany, Malaysia, Peru, Poland, Singapore, Spain, Turkey, Ukraine, and the USA. The two-factor structure was tested for invariance across Bangladesh, Ethiopia, Greece, India, Kenya, Morocco, Nepal, Philippines, Taiwan, Thailand, Namibia English, Namibia Oshiwambo, Namibia Khoekhoegowab

nation sample. For the one-factor countries, ANOVA with LSD post hoc tests indicated a significant difference only between Malaysia and Germany ($p = .003$). For the two-factor group, a significant difference in RA was found between Morocco and four nations: Greece, Bangladesh, Nepal, and Taiwan ($p < .01$). A significant difference in IR was found between Morocco and six nations: India, Ethiopia, Kenya, Namibia, Thailand (all $p < .001$), and the Philippines ($p < .01$). Interestingly, it can be seen here that although two factors are not highly correlated in these samples, they generally move together in a kind of rank order consistency, providing a general picture of the comparative religiosity between the nations.

Colors are used to group countries by region or into the category of “Western industrialized” (European and North American). It can be seen graphically here that the contexts in which prior research suggests we should expect lower religiosity, for example the industrialized West, belong to the one-factor group, while the contexts where the highest religiosity is expected belong to the two-factor group, with comparison between the two groups restricted due to model differences. To tentatively assess this, a post hoc comparison was made using the average raw score across five, two, or three items for the total score and two subscales, placing them on the same scale in part **c** of Fig. 3. Here it can be seen that the average raw item response for the two-factor countries, in particular for the African and Asian countries, was typically higher than in the one-factor set, in particular with regard to the Western and South

American countries. The statistical significance of these comparisons, however, cannot reliably be tested.

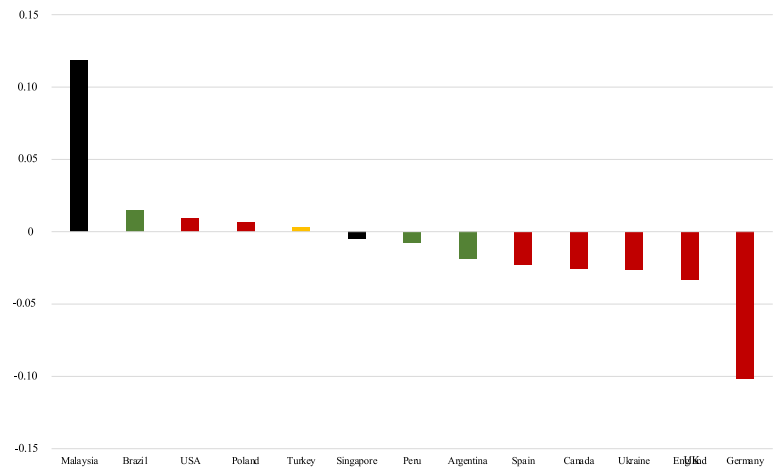
To further explore differences in variability within countries on the raw scores, a graphic comparison of the standard deviation for the average DUREL combined items for each country is displayed in Fig. 4. A group of seven “high-variability” countries are noticeable at the high end of the graph. These include the USA, Canada, Turkey, Brazil, Singapore, Poland, and the UK. The lowest variability was seen in the countries with the lowest and least reliable scores, Japan and China, followed by several of the countries with the highest averages.

Association with gender

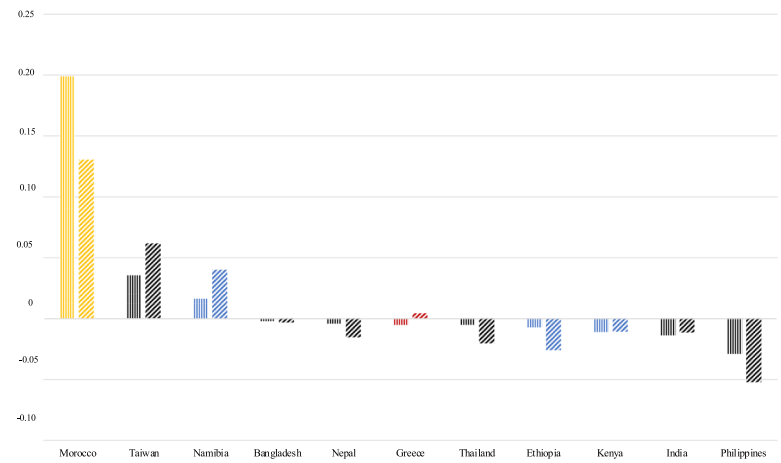
In Fig. 5, gender differences in latent DUREL scores are displayed graphically. Full details of the differences and ANOVA tests are in Supplemental Table S6. For the one-factor group, women had significantly higher total scores than men in 5 of the 13 samples (Poland, Ukraine, Spain, Brazil, Peru), and in no cases did men have higher scores than women. In the two-factor group, women reported significantly more religious activity in Namibia and Nepal, but men in Thailand. Women reported higher Intrinsic Religiosity in Kenya, Namibia, Morocco, and Nepal, and men only in Thailand. In 14 countries (Argentina, UK, USA, Canada, Germany, Greece, Turkey, Bangladesh, India, Taiwan, Singapore, Malaysia, Philippines, Ethiopia), no significant gender differences were found.

(See figure on next page.)

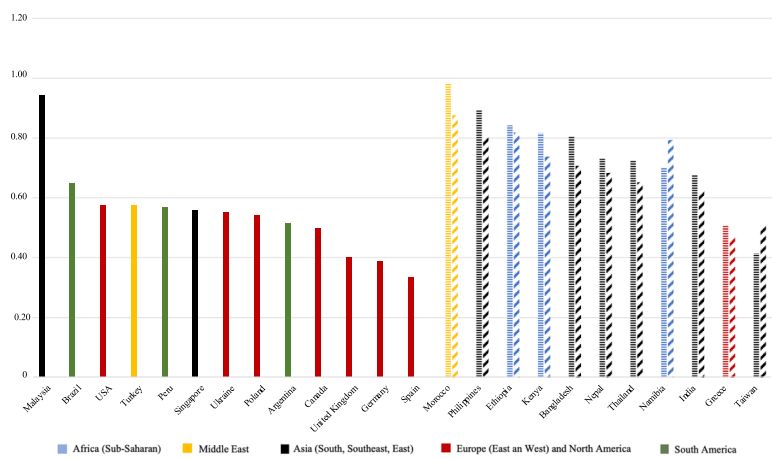
Fig. 3 Mean DUREL latent scores for **a** “one-factor” nations, from highest to lowest total score, and **b** “two-factor” nations, from highest to lowest RA. *Note.* In part **a**, a mean difference at $p < .01$ was found between Malaysia and Germany. In part **b**, a mean difference in RA at $p < .01$ was found between Morocco and four nations: Greece, Bangladesh, Nepal, and Taiwan. A mean difference in IR at $p < .001$ was found between Morocco and six nations: India, Ethiopia, Kenya, Namibia, Thailand, and the Philippines ($p < .01$). Part **c** uses raw scores, with scales divided by the number of items to create a comparable metric of an “average item response” on the DUREL’s 0–5 Likert scale



a) Highest to Lowest Total Score for Nations where One-Factor is Best Fit



b) Highest to Lowest Subscale Scores for Nations where Two-Factors is Best Fit



c) Post-Hoc Comparison of All Samples, Using Average Item-Level Raw Response

Fig. 3 (See legend on previous page.)

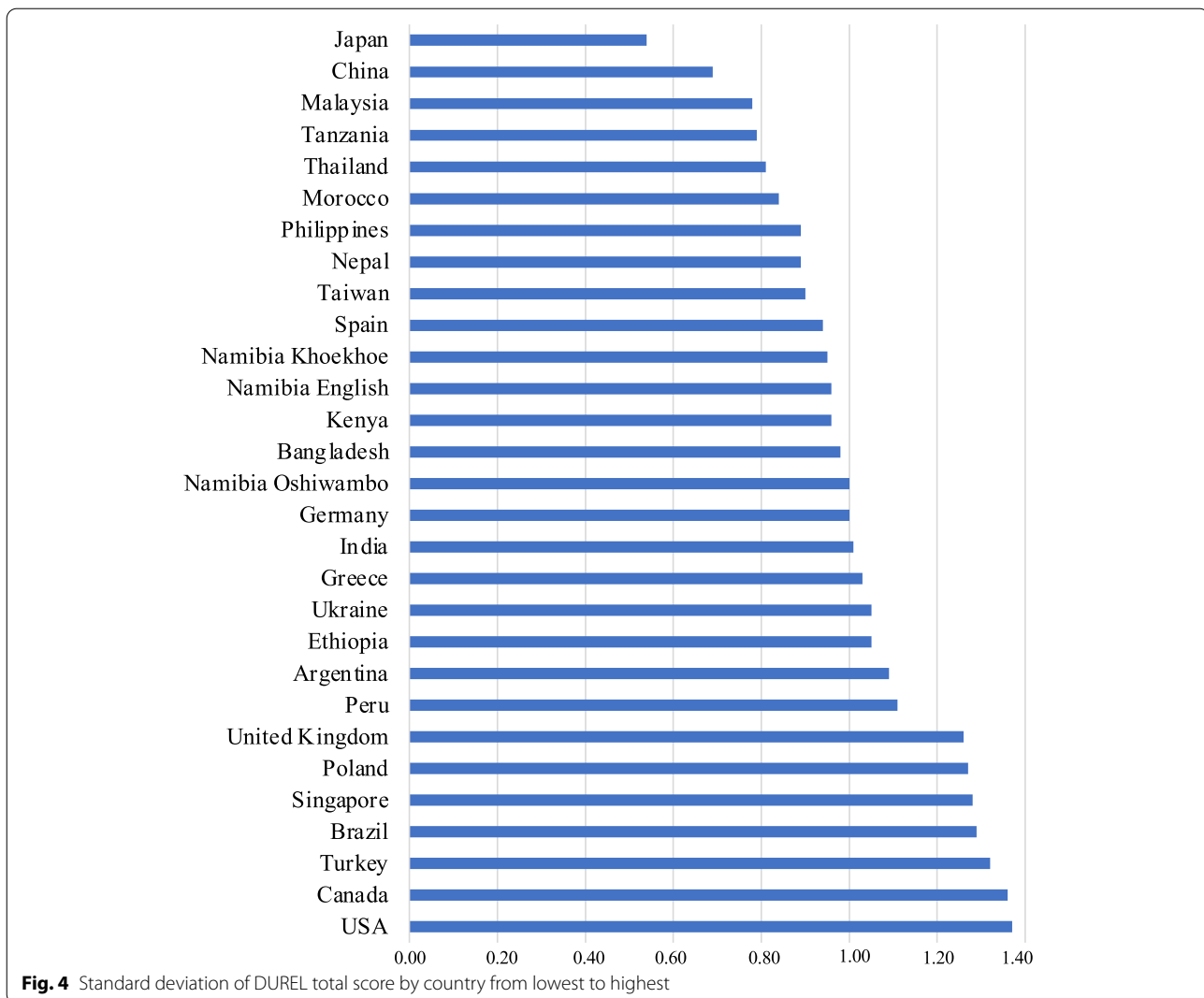


Fig. 4 Standard deviation of DUREL total score by country from lowest to highest

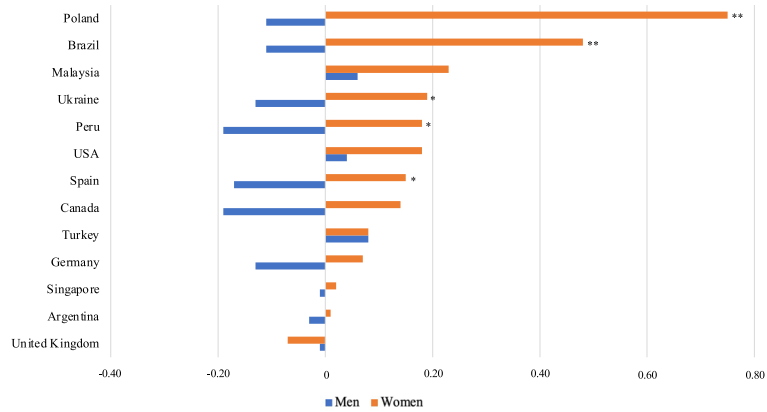
Discussion

In Study 2, we assessed the psychometric and structural properties of the DUREL in a global sample of 26 countries. This included tests in 18 total languages, in many of which the DUREL was not previously available, and the first tests of cross-cultural measurement invariance for the DUREL. All translations (as well as 13 others created for the project not tested due to small samples sizes) are freely available for research use at <https://psychometriglossia.uoregon.edu/>.

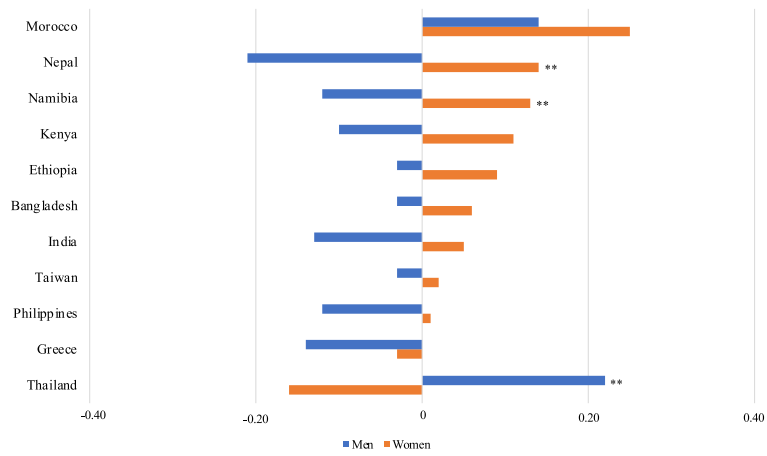
In 10 of the 26 countries, most of which are “majority world” contexts, results indicated better fit for a two-factor model with subscales of IR and of RA (combining the single-item components for organizational and non-organizational activity) than for a one-factor, total score model. In 13 countries, half of them Western industrialized contexts, a one-factor structure fits better; the

difference between the two aspects was not large enough to justify the use of subscales. Three countries were excluded from analysis for lack of variance, where religiosity is consistently very low, in China and Japan, and very high, in Tanzania.

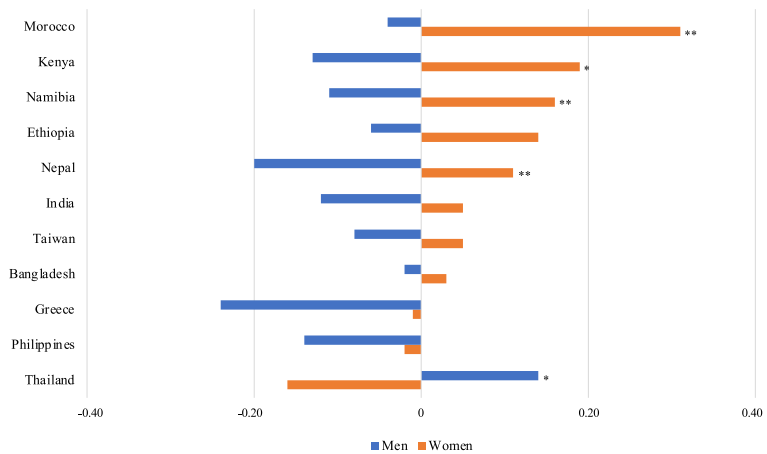
Tests of a one-dimensional model in 13 countries and for a two-dimensional model in 13 groups (including three Namibian samples) confirmed scalar-level measurement invariance for both sets. This means that latent means and variances for the DUREL can be meaningfully compared across countries within these sets, facilitating cross-cultural research in some contexts. Regional contrasts can only be drawn within the two categories, however, posing a barrier to many potentially interesting comparisons. This difficulty underlines the challenges of cross-cultural research. On average, religious engagement is certainly higher in African than in Western countries, but the extent of that difference, and the



(a) Total score, one-factor countries only



(b) Religious Activity, two-factors countries only



(c) Intrinsic Religiosity, two-factors countries only

Fig. 5 Gender differences for DUREL latent scores for one-factor and two-factor groups. Note. Countries in each graph are ordered from highest to lowest score for women. Latent scores are centered for the dataset and are positive or negative based on their variation from the sample mean. * $p < .01$; ** $p < .001$

comparison of individuals and subgroups between these contexts cannot reliably be estimated with the DUREL: the South American and Western countries are mainly in the one-factor group, while African countries are in the two-factor group; the Middle Eastern countries are split, and six of eight Asian countries are in the two-factor group. Given the acceptable (or close to it) fit of the one-factor model even in the countries where two factors fit better, a post hoc test of measurement invariance on the one-factor model across all 27 samples was added and is included in [supplemental materials](#).

With the exception of Malaysia, the one-factor group appeared to include the less religious countries. To tentatively assess this, a post hoc comparison was made to display the average item-level raw score for the total score or two subscales together. We see here that it is indeed primarily countries with lower average religiosity that fit a one-factor model. In general, in less religious contexts, the difference between religious activities and personal religious sentiment seems to be less pronounced. Individuals who identify as religious within these contexts typically report both, whereas, in contexts that are more religious overall, the two-factor structure is more useful, with individuals in the society expressing more varied patterns in or types of engagement. Malaysia is the exception, perhaps due to rates of religiosity so high that the scale lacks differentiation. This may be similar to the lack of reliability for the measure in the other most extreme contexts, Tanzania, China, and Japan. It appears that the DUREL works most reliably in contexts with enough variability that people can meaningfully think about and report their individual differences on these constructs.

To compare regions and religious contexts (see Supplemental Table S1 for national religious profiles) as best we can given these limitations, we can say that the sub-Saharan African countries in the sample, all of which are predominantly Christian, had higher DUREL scores than the similarly Christian Western industrialized and South American countries; that two Muslim countries, Morocco and Malaysia, seem to have the highest DUREL scores in the sample, and the Asian region was the most diverse, including both the highest (Malaysia) and the lowest scorers (Taiwan, China, Japan). Within-country variation in religious involvement appears to be lowest in the countries with both the lowest and the highest average scores, in some cases preventing reliable measurement with the DUREL. Particularly high variability in scores was seen in the industrialized Western countries, as well as in Turkey, Brazil, and Singapore, suggesting a great deal of diversity in engagement in some contexts.

We found women to be more engaged with religion than men in a number of nations, the one exception being Buddhist Thailand, where men report both more

Religious Activity and Intrinsic Religiosity. In more than half the countries sampled, however, including all the Western ones but Poland, all the Asian ones but Thailand, plus Argentina, Turkey, and Ethiopia, no significant gender differences were found.

Overall discussion

Religiousness, spirituality, and related attributes may have more importance in the study of psychology than has previously been assumed (Saucier, 2019). Measures of religiousness show cross-time stability as high or higher than for personality traits (Saucier, 2008; Saucier & Skrzypińska, 2006), meaning that these attributes fit the definition of personality as it is typically defined, as stable dispositions that reflect recurrent patterns of thinking, behavior, and emotions (e.g., as in Funder, 2013). Furthermore, religiousness is inarguably central in the study of psychology across cultures, capturing some of the variables that vary the most dramatically across nations (e.g., Saucier et al., 2015; Vaclair & Fischer, 2011; Remizova et al., 2022). In this study, the validity of the Duke University Religious Index was assessed in 30 cultural contexts and 20 languages, including those in which diverse religions predominate or are mixed, in order to facilitate the efficient assessment of this important construct. We were able to provide evidence of strong cross-cultural measurement invariance for the DUREL, and to recommend the best use of the index, either as a total score or with two subscales, in different contexts. Together, these studies provide the first rigorous assessments of measurement invariance for the DUREL, and thus allow us to draw useful comparisons of variation in relations involvement across national and religious contexts. They also provide 20 new translations of the DUREL to the scientific community, including the first into African languages.

The DUREL was designed to distinguish between three aspects of religiosity: Organizational Religious Activity, Non-Organizational Religious Activity, and Intrinsic Religiosity (Koenig & Büssing, 2010). Though the original authors recommended against the use of an aggregate total score when studying health outcomes, later studies have used it this way and demonstrated psychometric support for doing so. We considered this empirically. Given the high correlations between the components of the DUREL in many samples, we found that the use of a total score is the optimal approach in about half of the contexts we studied. In the other half of the nation samples, we found a two-factor model separating Intrinsic Religiosity from the two components related to Religious Activity (combined) to be better justified. The differences between the subscales in these contexts were not dramatic, never orthogonal or negatively correlated, but the fit for the two-factor model was appreciably better

than for the one-factor model. Future work might more thoroughly explore the potential for establishing measurement invariance across regions and samples, perhaps by making modifications (e.g., dropping an item, releasing parameters). Future work might also explore the reliability of item-level analysis that separate organized from private religious activity where useful.

Our results replicate international surveys of religiosity in finding DUREL scores to generally be higher in many majority world contexts. Western countries include more within-nation variation in scores, indicating the utility of this construct as an individual difference, and the potential importance of accounting for this difference in studies of other psychological phenomena. In Sub-Saharan Africa and in Muslim countries, the pervasiveness of religious engagement and feelings suggests their relevance to psychology in other regards, for example, structuring local discourses around mental health, healing, motivation, and values, and even personality description (e.g., Thalmayer, Job, et al., 2021). In Asia, dramatic intraregional differences demonstrate the importance of national experiences for psychological variables. While Asia is sometimes be thought of by Western researchers as monolithically “collectivist,” these results highlight very important differences in values, behaviors, and beliefs in a large and highly varied region.

The issue of appropriate cross-cultural assessment is also relevant for understanding the association of religiosity to other psychological and sociodemographic indicators. In a large prior meta-analysis of survey studies that included personality and religiosity variables, gender was not found to be a consistent predictor of religiosity (Saroglou, 2010), but our results suggest that this may be truer in Western countries. Cultural context and the type of religious involvement are important in unpacking this question, as seen in the World Values Survey (Sullins, 2006; Vardy et al., 2022). Sullins (2006) reported that the organizational structure of religious institutions can explain why women report less religious involvement in Muslim and Orthodox Jewish contexts, where they are excluded from many practices. This is likely also the explanation for our result from Buddhist Thailand, where living as monk for some months is a rite of passage to adulthood for most young men, but not for women (e.g., Assanangkornchai et al., 2002). Outside places with exclusionary structures, however, our results match those of many prior studies (e.g., Argyle & Beit-Hallahmi, 2013; Sullins, 2006) in finding that women report higher religious involvement in the countries where a difference is observed. The reasons for this are likely multi-faceted, encompassing aspects of the religious institutions and how participation is constructed, of the presence or

lack of other social and economic opportunities, of survey methodology and response patterns (Vardy et al., 2022; Sullins, 2006), and even of the local conceptualization of God (Vardy et al., 2022).

In the Namibian data, we were able to explore the association of the DUREL with well-being, physical health, age, and socioeconomic variables. Prior research suggests small positive associations for religiousness with life satisfaction (e.g., Berthold & Ruch, 2014), physical health and longevity (Chida et al., 2009; Koenig & Larson, 2001; McCullough et al., 2000; Seybold & Hill, 2001). In Namibia, we also found small positive associations. In future it would be useful to assess these associations in other contexts to determine if cultural conditions (e.g., type of religion, degree of religious diversity, or variation in religiousness within the country) play a role in the positive outcomes of religious involvement. Another important future direction could be to explore the role of religion as an effective protective factor in under-resourced contexts. Recent findings show that in the poorer countries where national religiosity is higher, low socioeconomic status has less negative impact on well-being than in richer countries (Berkessel et al., 2021) and that traumatic experiences are less likely to develop into ongoing post-traumatic stress disorder (Dückers et al., 2016). One hypothesis is that widespread religious engagement, including the strong communal bonds it nurtures, helps increase resilience.

Despite stereotypes in the West that link religiousness with lower education, embodied in Karl Marx’s quote that “religion is the opium of the people,” the association of religiousness with socioeconomic status and education appears to have been little explored. People with lower education have been seen to be more religious in Iran (Hafizi et al., 2014) and Portugal (Lucchetti et al., 2012) and in a 40-nation study that primarily sampled richer and Western countries (Schwadel, 2015). In Namibia, we found a different picture, with more educated and more employed individuals reporting more rather than less religious involvement, where there were any differences. This suggests the importance of considering cultural context for this association. Given the importance of religion to Namibians and to Africans generally, understanding the ways that it interacts with social and economic factors and opportunities would be important to understand more fully.

A limitation of the Survey of World Views data is that it consists only of college students. This is useful for isolating the effect of national affiliation from that of education level and age, but it limits generalizability to other members of the countries studied. Future work should assess the DUREL among other age groups. Additionally, while

this dataset includes many variables related to attitudes, beliefs, and values, it did not include the kind of health, well-being, or other life outcome variables that would allow for assessing the practical significance of religiosity in people's lives. Likewise, future work should explore the relation of religiosity with other stable personality characteristics across cultural contexts, to elucidate the ways this tendency complements or overlaps with commonly measured traits.

Another possible limitation stems from the applicability of the DUREL in contexts of religious traditions that do not focus on the divine, such as Buddhism, where the item "In my life, I experience the presence of the Divine (i.e., God)" could be less applicable. However, we note that we saw good measurement properties in both predominantly Buddhist (Thailand) and predominantly Hindu (India, Nepal) samples in this study. Future work could include additional, more diverse indicators of religiosity such as practices, orientation, beliefs, and community, following recent work by Remizova et al. (2022).

The importance of religious involvement to psychology has been underlined by recent work in the large samples of the World Values Survey. For example, Schulz et al. (2019) argue that the family policies of the Western Church altered traditional collectivistic kinship networks, setting in motion the structural and lifestyle changes that in turn shaped the individualistic, analytic, less conforming and more impersonally prosocial mindset that now predominates in Western cultures. In the same data, White et al. (2020) document cultural similarity between co-religionists and also between non-religious individuals both within and across countries, showing how denominations reflect shared historical descent, and indicating the cultural signature of religion on human psychology and on cultural boundaries. We hope that evidence for the cross-cultural validity and utility of the DUREL, with its straightforward assessment of religious involvement and sentiment, will facilitate the inclusion of this important construct in future work in both personality and cultural psychology, to aid the integration of these areas of study, and to account for this variable of particular importance in the majority world.

Conclusion

We assessed the Duke University Religious Index in 30 total contexts and 20 languages, providing evidence of strong cross-cultural measurement invariance despite the representation of many different religions in these samples. Together, these studies allow us to compare religious involvement globally, with the first rigorous assessments of measurement invariance, and to offer

19 new translations to the scientific community, including the first published translations of the inventory into African languages. Religiousness and spirituality are stable individual traits, thus belonging to the study of personality. They also capture (and arguably lead to) some of the largest measurable psychological differences between nations, indicating their importance in understanding how cultural context shapes the mind, and the need to take these constructs into account in both applied and research psychology. We hope that evidence for the cross-cultural validity and utility of the DUREL will facilitate the increased inclusion of religious sentiment and activity in both personality and cultural psychology, additionally supporting the integration of these areas of study.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s42409-022-00041-2>.

Additional file 1: Supplemental Table S1. National Religious Profiles of All Nations in Studies 1 and 2. **Supplemental Table S2.** Sample Characteristics by Language Group for Study 1. **Supplemental Table S3.** Results of the Assessment of Optimal EFA Structure for Study 1, Three Language Groups. **Supplemental Table S4.** Descriptive Statistics, and Group Differences for DUREL Scales (latent scores). **Supplemental Table S5.** One Versus Two Factors in Three Languages in Namibia, Using Scaled Fit Measures. **Supplemental Table S6.** Mean (and Standard Deviation) by Gender for DUREL Total Score and Subscales by Nation.

Acknowledgements

For Study 1, Paulus Mwetulundila and Matilda Maletzky recruited and helped coordinate the English and Oshiwambo language teams, and Sylvanus Job and Matilda Maletzky the Khoekhoegowab team. For Oshiwambo, translators included Petrus Mbenzi, Rauna Ndinoshio, Lea O. N. Haidula, and Emily Kapena Kambonde. Interviewers were Esther Muhepa, Regina Mwanyangapo, Taimi Amutse, Julius Ndikwetepo, Selma Inghepa, Aina Sakaria, Benjamin Shikesho, Vanyenga Mandume, and Mariana Mbeeli. In Khoekhoegowab translators included Jeremiah Goeieman, Christine Luipert, and Sylvanus Job; interviewers were Morina Angula, Elverardo JAwases, Eljakim "Baby" Doeseb, Kuti |Hoxobeb, Fritz Isak Dirkse, Marialda Garises, Jeremiah Goeieman, Valerie Isaaks, Lorensia Kafuro, and Venessa Uiras. In English, data collection additionally included Vicky Shipale, Shasheeda Mberira, Ndamono Shikoyeni, Lisilzenda Garises, and Violar Sinami. Gayoung Son of the University of Basel helped coordinate data collection.

For Study 2, we thank Gerard Saucier of the University of Oregon for making the Survey of World Views data available to us.

Authors' contributions

AGT, CT, ENS, and SLR contributed to conception and design of the overall study. AGT obtained funding. For Study 1, AGT and ENS handled ethical review, study permitting, translation of the inventories, training research assistants and monitoring data collection. SLR, CT, and AGT planned the statistical analyses. SLR performed analyses in R, CT descriptive analyses in SPSS. CT and AGT wrote the initial draft of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

Funding

Work on this article was supported by Swiss National Science Foundation grant 10001C_179458.

Availability of data and materials

A pre-registered analysis plan and study materials are available on the Open Science Framework (https://osf.io/y8d4z/?view_only=347d245ea9544b0)

9a4111a8650059af4; (https://osf.io/6d8gs/?view_only=a32710f4784e447f890aa077ff6b5bc9) [blinded for peer review]. Translations of the DUREL into 31 languages are available at <https://psychometriglossia.uoregon.edu/>.

Declarations

Ethics approval and consent to participate

Ethical review of the study plan was made by the University Research Ethics Commission of the University of Namibia in 2019. Data was collected in August 2019. Written informed consent was obtained from all participants before they completed the survey.

Consent for publication

All authors have seen the manuscript and consent to its publication.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Institute of Psychology, University of Lausanne, Lausanne, Vaud, Switzerland. ²Research Group Work & Organisational Psychology – KU Leuven, Leuven, Belgium. ³Department of Psychology and Social Work, University of Namibia, Windhoek, Namibia. ⁴Department of Psychology, University of Zürich, Binzmühlestrasse, 8050 Zürich, Switzerland.

Received: 26 May 2022 Accepted: 16 September 2022

Published online: 10 October 2022

References

- Argyle, M., & Beit-Hallahmi, B. (2013). *The social psychology of religion*. Routledge. <https://doi.org/10.4324/9780203794692>.
- Assanangkornchai, S., Conigrave, K. M., & Saunders, J. B. (2002). Religious beliefs and practice, and alcohol use in Thai men. *Alcohol and Alcoholism*, *37*(2), 193–197. <https://doi.org/10.1093/alcac/37.2.193>.
- Bentler, P. M. (1995). *EQS: Structural equations program manual*. Multivariate Software.
- Berkessel, J. B., Gebauer, J. E., Joshanloo, M., Bleidorn, W., Rentfrow, P. J., Potter, J., & Gosling, S. D. (2021). National religiosity eases the psychological burden of poverty. *Proceedings of the National Academy of Sciences*, *118*(39). <https://doi.org/10.1073/pnas.2103913118>.
- Berthold, A., & Ruch, W. (2014). Satisfaction with life and character strengths of non-religious and religious people: It's practicing one's religion that makes the difference. *Frontiers in Psychology*, *5*, 1–9. <https://doi.org/10.3389/fpsyg.2014.00876>.
- Blunch, N. (2012). *Introduction to structural equation modeling using IBM SPSS statistics and AMOS*. Sage.
- Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research*, *6*(2), 229–249. <https://doi.org/10.1086/681607>.
- Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. <https://doi.org/10.1080/15305051003637306>.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. <https://doi.org/10.1177/0049124108314720>.
- Chen, H., Zhizhong, W., Philipps, M., Sun, Y., & Cheng, H. (2014). Internal consistency and test-retest reliability of the Chinese version of the 5-item Duke University Religion Index. *Shanghai Archives of Psychiatry*, *26*(5), 300–309. <https://doi.org/10.11919/j.issn.1002-0829.214088>.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Cheung, G. W., & Rensvold, R. B. (2009). Structural equation modeling evaluating goodness-of-fit indexes for testing measurement invariance. *Taylor & Francis*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5.
- Chida, Y., Steptoe, A., & Powell, L. (2009). Religiosity/spirituality and mortality. *Psychotherapy and Psychosomatics*, *78*(2), 81–90. <https://doi.org/10.1159/000190791>.
- De Berardis, D., Olivieri, L., Giardini, A., Rapini, G., Serroni, N., Fornaro, M., ... Di Giannantonio, M. (2020). Religiosity as a protective factor against suicide ideation in subjects with Major Depression: preliminary results of an exploratory study in the “real world” clinical practice. *Evidence-based Psychiatric Care*, *6*, 107–111. <https://doi.org/10.1097/NMD.0b013e31826b6d05>.
- DeSalvo, K. B., Bloser, N., Reynolds, K., He, J., & Muntner, P. (2006). Mortality prediction with a single general self-rated health question: a meta-analysis. *Journal of General Internal Medicine*, *21*(3), 267–275. <https://doi.org/10.1111/j.1525-1497.2005.00291.x>.
- DeSalvo, K. B., & Muntner, P. (2011). Discordance between physician and patient self-rated health and all-cause mortality. *The Ochsner Journal*, *11*(3), 232–240.
- Diener, E., & Emmons, R. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13.
- Dobrowolska, B., Jurek, K., Pilewska-Kozak, A. B., Pawlikowski, J., Drozd, M., & Koenig, H. (2016). Validation of the Polish version of the Duke University Religion Index (PolDUREL). *Polish Archives of Internal Medicine*, *126*(12), 1005–1008. <https://doi.org/10.20452/pamw.3721>.
- Dückers, M. L., Alisic, E., & Brewin, C. R. (2016). A vulnerability paradox in the cross-national prevalence of post-traumatic stress disorder. *The British Journal of Psychiatry*, *209*(4), 300–305. <https://doi.org/10.1192/bjp.bp.115.176628>.
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, *10*. <https://doi.org/10.3389/fpsyg.2019.01507>.
- Frydman, J. (2011). A critical analysis of Namibia's English-only language policy. In *Selected Proceedings of the 40th Annual Conference on African Linguistics*, (pp. 178–189).
- Funder, D. C. (2013). *The personality puzzle*, (6th ed.,). W.W. Norton.
- Gonzales, L., Koci, A., Gee, R. M., Noji, A., Glaser, D., Marsh, A. K., ... Al Sabei, S. (2015). Caring for women globally: Psychometric testing of two instruments translated into five languages for use in cardiovascular recovery. *International Journal of Nursing Practice*, *21*, 27–37. <https://doi.org/10.1111/ijn.12393>.
- Hafizi, S., Tabatabaei, D., & Koenig, H. G. (2014). Borderline personality disorder and religion: A perspective from a Muslim country. *Iranian Journal of Psychiatry*, *9*(3), 137–141.
- Hall, D. E., Meador, K. G., & Koenig, H. G. (2008). Measuring religiousness in health research: review and critique. *Journal of Religion and Health*, *47*(2), 134–163.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Jamshidian, M., & Jalal, S. (2010). Tests of homoscedasticity, normality, and missing completely at random for incomplete multivariate data. *Psychometrika*, *75*, 649–674. <https://doi.org/10.1007/s11336-010-9175-3>.
- Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R Package for Testing Homoscedasticity, Multivariate Normality, and Missing Completely at Random (MCAR). *Journal of Statistical Software*, *56*(6), 1–31 <http://www.jstatsoft.org/v56/i06/>.
- Jones, J. (2021). *U.S. Church Membership Falls Below Majority for First Time*. Retrieved from: <https://news.gallup.com/poll/341963/church-membership-falls-below-majority-first-time.aspx>.
- Kagitcibasi, C. (2002). A model of family change in cultural context. *Online Readings in Psychology and Culture*, *6*(3), 1–11. <https://doi.org/10.9707/2307-0919.1059>.
- Kenny, D. A. (2016). *Multiple Latent Variable Models: Confirmatory Factor Analysis*. Retrieved from: davidakenny.net/cm/mfactor.htm.
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. <https://doi.org/10.1177/0049124114543236>.
- Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 524–544. <https://doi.org/10.1080/10705511.2017.1304822>.

- Koenig, H. G., & Büssing, A. (2010). The Duke University Religion Index (DUREL): A five-item measure for use in epidemiological studies. *Religions*, 1(1), 78–85. <https://doi.org/10.3390/rel1010078>.
- Koenig, H. G., & Larson, D. B. (2001). Religion and mental health: Evidence for an association. *International Review of Psychiatry*, 13(2), 67–78. <https://doi.org/10.1080/09540260124661>.
- Lace, J. W., & Handal, P. J. (2018). Confirming the tripartite structure of the Duke University Religion Index: A methodological approach. *Journal of Religion and Health*, 57(2), 704–716. <https://doi.org/10.1007/s10943-017-0556-0>.
- Lee, E., & Baumann, K. (2013). German psychiatrists' observation and interpretation of religiosity/spirituality. *Evidence-based Complementary and Alternative Medicine*, 1–8. <https://doi.org/10.1155/2013/280168>.
- Lipsmeyer, M. E. (1984). *The measurement of religiosity and its relationship to mental health/impairment*. St. Louis: Saint Louis University.
- Lucchetti, G., Lucchetti, A. L. G., Peres, M. F., Leão, F. C., Moreira-Almeida, A., & Koenig, H. G. (2012). Validation of the Duke Religion Index: DUREL (Portuguese Version). *Journal of Religion and Health*, 51(2), 579–586. <https://doi.org/10.1007/s10943-010-9429-5>.
- Martins, H., Caldeira, S., Domingues, T. D., Vieira, M., & Koenig, H. G. (2021). Validation of the Duke University Religion Index (DUREL) in Portuguese Cancer Patients Undergoing Chemotherapy. *Journal of Religion and Health*, 1–14. <https://doi.org/10.1007/s10943-020-01143-z>.
- McCullough, M. E., Hoyt, W. T., Larson, D. B., Koenig, H. G., & Thoresen, C. (2000). Religious involvement and mortality: a meta-analytic review. *Health Psychology*, 19(3), 211–222. <https://doi.org/10.1037/0278-6133.19.3.211>.
- McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, 6(1), 97–103. <https://doi.org/10.1007/BF01908590>.
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. <https://doi.org/10.1037/1082-989X.7.1.64>.
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Non-centrality and goodness of fit. *Psychological Bulletin*, 107(2), 247–255. <https://doi.org/10.1037/0033-2909.107.2.247>.
- Nurasikin, M. S., Aini, A., Ahmad Adlan, A. S., & Ng, C. G. (2010). Validity and reliability of the Malay version of Duke University Religion Index (DUREL-M) among a group of nursing student. *Malaysian Journal of Psychiatry*, 19(2), 68–72.
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers*, 32(3), 396–402. <https://doi.org/10.3758/bf03200807>.
- Pengpid, S., Peltzer, K., & Ahsan, G. U. (2015). Risk of eating disorders among university students in Bangladesh. *International Journal of Adolescent Medicine and Health*, 27(1), 93–100. <https://doi.org/10.1515/ijamh-2014-0013>.
- Pew Research Center (2012). *The Global Religious Landscape*<https://assets.pewresearch.org/wp-content/uploads/sites/11/2014/01/global-religion-full.pdf>.
- Pew Research Center (2018). *The age gap in religion around the world*<https://www.pewforum.org/wp-content/uploads/sites/7/2018/06/ReligiousCommitment-FULL-WEB.pdf>.
- Plante, T. G., Vallaeys, C. L., Sherman, A. C., & Wallston, K. A. (2002). The development of a brief version of the Santa Clara Strength of Religious Faith Questionnaire. *Pastoral Psychology*, 50(5), 359–368. <https://doi.org/10.1023/A:1014413720710>.
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. *Developmental Review*, 41(9), 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>.
- R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing<https://www.Rproject.org/>.
- Remizova, A., Rudnev, M., & Davidov, E. (2022). In Search of a Comparable Measure of Generalized Individual Religiosity in the World Values Survey. *Sociological Methods & Research*, 00491241221077239. <https://doi.org/10.1177/00491241221077239>.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://www.jstatsoft.org/v48/i02/>.
- Saffari, M., Zeidi, I. M., Pakpour, A. H., & Koenig, H. G. (2013). Psychometric properties of the Persian version of the Duke University Religion Index (DUREL): A study on Muslims. *Journal of Religion and Health*, 52(2), 631–641. <https://doi.org/10.1007/s10943-012-9639-0>.
- Saroglou, V. (2010). Religiosity as a cultural adaptation of basic traits: A five-factor model perspective. *Personality and Social Psychology Review*, 14(1), 108–125. <https://doi.org/10.1177/1088868309352322>.
- Saucier, G. (2008). Measures of the personality factors found recurrently in human lexicons. In G. J. Boyle, G. Matthews, & D. Saklofske (Eds.), *Handbook of personality theory and testing: Vol. 2, Personality measurement and assessment*, (pp. 29–54). SAGE.
- Saucier, G. (2019). Is religiosity a form of variation in personality, or in culture, or neither? Conceptual issues and empirical indications. *Archive for the Psychology of Religion*, 41(3), 216–223. <https://doi.org/10.1177/0084672419894677>.
- Saucier, G., Kenner, J., Iurino, K., Bou Malham, P., Chen, Z., Thalmayer, A. G., ... Altschul, C. (2015). Cross-cultural differences in a global "Survey of World Views". *Journal of Cross-Cultural Psychology*, 46(1), 53–70. <https://doi.org/10.1177/0022022114551791>.
- Saucier, G., & Skrzypięńska, K. (2006). Spiritual but not religious? Evidence for two independent dispositions. *Journal of Personality*, 74, 1257–1292. <https://doi.org/10.1111/j.1467-6494.2006.00409.x>.
- Savalei, V., & Rhemtulla, M. (2012). On obtaining estimates of the fraction of missing information from full information maximum likelihood. *Structural Equation Modeling: A Multidisciplinary Journal*, 19(3), 477–494.
- Schulz, J. F., Bahrami-Rad, D., Beauchamp, J. P., & Henrich, J. (2019). The Church, intensive kinship, and global psychological variation. *Science*, 366(6466), eaau5141. <https://doi.org/10.1126/science.aau5141>.
- Schwadel, P. (2015). Explaining cross-national variation in the effect of higher education on religiosity. *Journal for the Scientific Study of Religion*, 54(2), 402–418. <https://doi.org/10.1111/jssr.12187>.
- Seybold, K. S., & Hill, P. C. (2001). The role of religion and spirituality in mental and physical health. *Current Directions in Psychological Science*, 10(1), 21–24. <https://doi.org/10.1111/1467-8721.00106>.
- Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 1–15. <https://doi.org/10.1080/10705511.2019.1611434>.
- Smith, T. B., McCullough, M. E., & Poll, J. (2003). Religiosity and depression: Evidence for a main effect and the moderating influence of stressful life events. *Psychological Bulletin*, 129(4), 614–636. <https://doi.org/10.1037/0033-2909.129.4.614>.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4.
- Storch, E. A., Roberti, J. W., Heidgerken, A. D., Storch, J. B., Lewin, A. B., Killiany, E. M., ... Geffken, G. R. (2004). The duke religion index: A psychometric investigation. *Pastoral Psychology*, 53(2), 175–181. <https://doi.org/10.1023/B:PASP.0000046828.94211.53>.
- Sullins, D. P. (2006). Gender and religion: Deconstructing universality, constructing complexity. *American Journal of Sociology*, 112(3), 838–880. <https://doi.org/10.1086/507852>.
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using M plus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal*, 27(1), 111–130. <https://doi.org/10.1080/10705511.2019.1602776>.
- Tarakeshwar, N., Stanton, J., & Pargament, K. I. (2003). Religion: An overlooked dimension in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, 34(4), 377–394. <https://doi.org/10.1177/0022022103034004001>.
- Thalmayer, A. G., Job, S., Shino, E. N., Robinson, S. L., & Saucier, G. (2021). #Usgu: A mixed-method lexical study of character description in Khoekhoe-gowab. *Journal of Personality and Social Psychology*, 121(6), 1258–1283. <https://psycnet.apa.org/doi/10.1037/pspp0000372>.
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129. <https://doi.org/10.1037/amp0000622>.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. <https://doi.org/10.1007/BF02291170>.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations

- for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>.
- Vardy, T., Moya, C., Placek, C. D., Apicella, C. L., Bolyanatz, A., Cohen, E., ... Atkinson, Q. D. (2022). The religiosity gender gap in 14 diverse societies. *Religion, Brain and Behavior*, 12(1-2), 18–37. <https://doi.org/10.1080/2153599X.2021.2006292>.
- Vauclair, C.-M., & Fischer, R. (2011). Do cultural values predict individuals' moral attitudes? A cross-cultural multilevel approach. *European Journal of Social Psychology*, 41(5), 645–657. <https://doi.org/10.1002/ejsp.794>.
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41, 321–327. <https://doi.org/10.1007/BF02293557>.
- Wang, Z., Rong, Y., & Koenig, H. G. (2014). Psychometric properties of a Chinese version of the Duke University Religion Index in college students and community residents in China. *Psychological Reports*, 115(2), 427–443. <https://doi.org/10.2466/08.17.PR0.115c19z8>.
- White, C., Muthukrishna, M., & Norenzayan, A. (2020). Worldwide evidence of cultural similarity among co-religionists within and across countries using the World Value Survey. *Proceedings of the National Academy of Sciences of the United States of America*, 118(37), 1–9. <https://doi.org/10.1073/pnas.2109650118>.
- 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. <https://doi.org/10.1007/s11336-016-9506-0>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

