Cross-cultural assessment and comparisons of risk tolerance across domains

Yiyun Shou[1] · Joel Olney[1] · Meng-Cheng Wang[2]

Abstract

Risk attitudes are known to play an important role in influencing one's behavior under conditions of uncertainty. To date, cultural influences on risk attitudes - beyond the effects they have on perceived risk - have not been well understood. Having a cross-culturally invariant measure of risk attitudes is a prerequisite for carrying out more in depth explorations in this area. The current study applied the domain-specific risk attitudes framework and focused on the Chinese and US cultural contexts. Using novel network analysis techniques, we explored domain-specific patterns of risk attitudes in Chinese and US community samples and we subsequently developed a version of the Multi-Domain Risk Tolerance scale (MDRT-EC) that had similar applicability in both samples. The MDRT-EC demonstrated excellent psychometric characteristics and achieved strong measurement invariance across both samples. The associations between MDRT-EC domain scales and criterion scales were also similar between the two samples, further indicating the measurement invariance of the MDRT-EC. Finally, we used the MDRT-EC to explore cultural differences in risk attitudes across domains and their predictive relations with a range of lifestyle behaviors.

Keywords Cross-cultural comparison · Domain specific risk attitudes · Risk tolerance · Measurement invariance

In much of the psychological literature, the presence of "risk" refers to the possibility of danger or negative outcomes in a given situation. Attitude toward risk, which influences how one reacts to and behaves under conditions of uncertainty, is therefore an important construct in many fields of psychology, including clinical, forensic, cognitive, and social psychology. The role culture and society play in determining risk attitudes have been the focus of research for decades now, and are important focal points in helping us to understand the construct of risk attitudes beyond the Euro-American cultural context. There have been increasing numbers of empirical studies investigating cultural differences in risk attitudes since the 1990s, primarily in the fields of business (e.g., consumer behaviors and management) and tourism. The present article will focus on the comparison of

risk attitudes across the Euro-American (Anglosphere) and East Asian (Sinosphere) cultural contexts.

Early studies employing behavioral tasks, such as monetary choice outcomes, demonstrated that student participants from East Asian (e.g., Chinese) backgrounds exhibited a greater risk-taking tendency than those from Euro-American backgrounds (Hsee & Weber, 1999; Weber & Hsee, 1998; Weber et al., 1998; Terpstra-Tong & Terpstra, 2013). Hsee and Weber (1999) proposed the cushion hypothesis to account for these findings, which posits that the collectivist nature of Chinese culture provides a buffer against the impact of financial loss on individuals, as the members of an individual's social network can share the burden of this loss (e.g., family and friends). On the other hand, two recent studies that measured risk attitudes via a range of participants' self-selected health and recreational activities suggested that Australian students exhibited greater risk tolerance than South Korean and Chinese students (Kim & Park, 2010; Park et al., 2015). Park et al. (2015) suggested that the cultural differences in these domains reflect that individuals in collectivist cultures, such as the East Asian context, were more likely to be constrained by mandated social customs or norms when facing uncertainty than those in individualist

✉ Yiyun Shou
   yiyun.shou@anu.edu.au

1   Research School of Psychology, The Australian National University, Canberra, ACT 2601, Australia

2   Department of Psychology, Guangzhou University, Guangzhou, China

cultures. These findings are backed up by research in the tourism literature that demonstrates that East Asian tourists are much less likely to engage in high-risk adventure activities than Euro-Americans (Pizam et al., 2004; Reisinger & Mavondo, 2006). This suggests that cultural differences in risk attitudes vary depending on the domain of risk in question as well as the nature of the tasks and measures.

Over the past two decades, research on risk attitudes has shifted towards a domain-specific framework (Figner & Weber, 2011; Weber & Johnson, 2009), with a primary focus on the behavioral tendency component of attitudes. The crux of a domain-specific understanding of risk attitudes is that individuals can vary their behavioral tendency across decision domains. For example, an individual who is willing to make a high-risk financial investment may not be willing to engage in actions with a potential for negative health consequences. According to the risk-return framework, this inconsistency is primarily due to individuals varying their perceptions of benefit and loss across domains, such that the subjective values (or the trade-off between benefit and loss) of their actions differ across decision domains. Weber and colleagues pioneered the theory and research of the domain-specific framework, and proposed the prominent Domain-Specific Risk-Taking (DOSPERT) scale to measure these domain-specific risk attitudes (Weber et al., 2002).

The domain-specific framework and DOSPERT have been fruitfully applied to understand how individuals may react differently across domains and how groups can differ in risk attitudes across domains. In terms of group differences, the perceived risk of a situation, including the negative consequences and their probabilities, can often be restricted by the decision environment. In this way, culture and society play a central role in shaping one's environment, and thus influence risk attitudes via perceived outcomes and probabilities (Weber & Hsee, 1998). Several studies have investigated domain-specific risk attitudes in East Asian samples using the DOSPERT. These studies have reported a number of findings similar to those using Euro-American samples. For example, Chinese participants scored higher on the social and recreational risk-taking domains than the ethical, financial and health domains (Cheung & Tao, 2013), and males scored higher on the ethical, financial and health domains of the DOSPERT than females (Du et al., 2014; Hu & Xie, 2012). In terms of the level of risk-taking, Du et al. (2014) reported that Chinese undergraduate students had higher scores on financial investment risk-taking, but similar scores on other domains compared to the average scores reported in Western samples.

Unfortunately, the large majority of previous studies that have investigated cultural differences in risk-taking across multiple domains did not assess the cross-cultural invariance of the measures they used. For example, Park et al. (2015)'s study used a measure in which participants only rated items that they felt were personally relevant. The items selected by Australian students could be different from those selected by South Korean or Chinese participants. Thus, a comparison in the mean scores between these two samples becomes problematic, as items were different for the different samples. On the other hand, scales such as DOSPERT have been developed within a Euro-American cultural context. As such, the validity and reliability of these measures may also not have been established in the target cultural groups. If a measure is not assessing the same construct across cultural groups, conclusions about comparisons of the mean scores can be biased and invalid. The current study aimed to explore cultural differences in risk attitudes, and, in particular, address the issue of measurement invariance.

Measuring Domain-Specific Risk Attitudes

DOSPERT, together with the revised version of the scale (Blais & Weber, 2006), has been one of the most widely used measures of domain-specific risk attitudes in the literature for the past twenty years. DOSPERT contains a range of behaviors (e.g., "Going down a ski run that is beyond your ability." from the recreational domain) across the following five life domains: ethical, financial, health, social and recreational. While previous research has demonstrated the validity of DOSPERT in terms of its correlations with external criterion variables, there is limited evidence for its structural validity and measurement invariance. A recent meta-analysis found that the internal consistency reliability of DOSPERT was notably low for certain domain scales, such as the social and health domains (Shou & Olney, 2020). Studies that used non-English speaking samples also reported significantly lower reliability in the ethical, social and health domains than studies that used English-speaking samples (Shou & Olney, 2020). Several of these studies reported remarkably low reliabilities, especially for the health and social domains in East Asian samples (Cheung & Tao, 2013; Cheung et al., 2016; Wang et al., 2017; Wichary et al., 2015). In terms of the factor structure and measurement invariance of the DOSPERT, few of these studies tested these properties across East Asian and Euro-American samples. The only study that did conduct these analyses reported poor model fit of a five-factor model in East Asian populations (Wu & Cheung, 2014). The lack of evidence concerning the measurement invariance of DOSPERT limits the assessment of risk attitudes for the purpose of cross-cultural studies, as the DOSPERT scores across the two cultural groups may not represent the same construct.

It was suggested that DOSPERT's lack of internal structural validity, especially in East Asian cultural groups, could be due to unclear domain definitions, item ambiguity, and the impact of heterogeneity in respondents' prior knowledge (Shou & Olney, 2020). First, the nature of the risk is not explicitly stated in most of the DOSPERT items. An item that is perceived as entailing risk in one particular domain from a certain cultural perspective may be perceived as entailing a different domain of risk altogether for a different cultural group. For example, having an affair with a married person, which is classified as an ethically risky behavior in DOSPERT, could be perceived as a socially risky behavior in the Chinese cultural context as such behavior could shame the whole family. Second, although DOSPERT has three subscales, including risk perception, benefit perception and behavioral intention, to assess risk attitudes relatively comprehensively, the majority of previous studies only use the behavioral intention scale of the DOSPERT (i.e., how likely it is one would engage in a particular behavior) for the purpose of test efficiency. The behavioral intention measurement approach might result in the measured traits being influenced by respondents' prior knowledge and experience of the behavioral situation. Items that are unfamiliar to participants could exhibit poor discriminant ability in assessing risk attitudes.

More recently, Shou and Olney (2021) proposed the Multi-Domain Risk Tolerance scale (MDRT) based on the DOSPERT framework to assess affective responses to domain-specific risk. The generation of the MDRT items aimed at addressing the measurement issues associated with the DOSPERT, including having clearer domain specification and lower item ambiguity (Shou & Olney, 2021). The authors have argued that the MDRT is a good alternative to other scales, such as DOSPERT, available in the literature for cross-cultural studies of risk attitudes. Firstly, the items specify risk information and thus can reduce the influence of participant familiarity and prior knowledge with the item contents. As such, participants are better informed of the type of risk in the item even if they have little experience about the behavior before using the scale. Second, the MDRT emphasises on affective responses to items (i.e., how pleasant one would feel toward a situation), which can reduce the influence of cultural differences in the feasibility of engaging in a behavior. For example, some adventure and recreational activities, such as white water rafting, are more common in countries such as Australia or US, but less accessible or known in China and Japan. Finally, clearer wording in terms of which domain of risk the item corresponds to can mitigate perceived domain ambiguity due to cultural differences.

The Current Study

The first aim of the current study was to develop and validate a Chinese version of the MDRT scale. The second aim was to investigate domain-specific risk attitudes between Chinese and US community samples using a version of the MDRT that has measurement equivalence between the two samples. Commonly, validation of an existing scale in a new cultural group is limited by the items that were selected and developed in the original cultural group. Items that function best for one group may not function equally well in another cultural group. There is limited space for adapting or modifying the scale to both accommodate a new cultural group and for cross-cultural comparisons. To enable cross-cultural comparisons of risk attitudes between the two samples, we started with the original item pool used in Shou and Olney (2021) and selected items to construct a version of the MDRT that was best suited to both the Chinese and US community samples. We name this version of the scale MDRT-EC.

We first apply Exploratory Graph Analysis for initial item selection to ensure that the selected items have similar clustering and connectedness in both samples. Subsequently, we ensure the measurement invariance of the joint version of the MDRT-EC across the two samples using multi-group confirmatory factor analysis. After establishing structural and metric measurement invariance, we compare the two groups in terms of the convergent pattern of the correlations between MDRT-EC and conceptually related constructs to strengthen the construct measurement invariance. We would examine the cultural differences in risk attitudes using the MDRT-EC only when scalar equivalence is established.

In addition, it was observed that lifestyle behaviors such as smoking, drinking and exercise were associated with tolerance of risk outside of, or beyond, the medical and health domain (Shou & Olney, 2021). For example, ethical and recreational risk tolerance were found to be positively correlated with alcohol consumption, while social risk tolerance positively correlated with engagement in exercise. Given that the legal and social norms around different lifestyle behaviors could be different between Chinese and US contexts, we explored cultural differences in how risk attitudes measured by MDRT-EC predict a range of lifestyle behaviors. Finally, a shortened version of the MDRT with the best performing items for the Chinese sample was also proposed.

Study 1

Study 1 aimed to develop the Chinese version of the original MDRT item pool using a Chinese-English bilingual sample. Participants were recruited via the online crowdsourcing platform Prolific and were required to be fluent in both

English and Chinese. Language proficiency was tested by both self-reported fluency as well as language test questions that featured an attention catch question in Chinese and an attention catch question in English. Sixty-two participants (31 males) completed the study and met the language requirements. The participants were aged between 18 and 42 ($M = 26.82$, $SD = 5.48$). All participants spoke English and Chinese as either their first language or fluently (24 spoke English as their first language and 32 spoke Chinese as their first language).

Participants completed the English and Chinese versions of the MDRT. The MDRT contains 52 items that cover six different domains: ethical, financial, health/medical, recreational-safety (recreational activities with health risks), social, and recreational (recreational activities with other risks) (See Shou & Olney, 2021 for a more in-depth explanation of the domains). Participants were asked to rate their feeling toward the situation described in each of the items. Each item was rated on a 7-point Likert scale from extremely unpleasant to extremely pleasant. The first author (the author of the English MDRT) translated the 52 items into Chinese. The third author (a native Chinese speaker and expert in psychology assessments) revised the translation.

Participants completed the study via the Qualtrics survey platform. The order of the two language versions was randomized. The project was approved by the Australian National University Human Research Ethics Committee (protocol number: 2017/915).

Paired correlations and t-tests were used to compare participants' ratings on the English and Chinese MDRT items. The results are presented in the online supplementary materials (Table S1). All items either had strong correlations between the two language versions ($r > 0.5$) or had no significant differences ($p < .05$) in the means between the language versions. This indicates reasonable convergence between the two language versions.

Study 2

Study 2 aimed to develop a joint version of the MDRT (MDRT-EC) for the cross-cultural comparison of risk attitudes. We expected that the network and factor structures of the MDRT-EC would be equivalent across Chinese and US samples. We also hypothesized that the MDRT-EC domain scales would significantly converge with conceptually relevant scales in a similar magnitude across both samples (bold in Table 3). Next, we explored the cultural differences in risk attitudes measured by the MDRT-EC and their associations with lifestyle behaviors. Finally, we proposed a shortened version of the MDRT with items that performed best for the Chinese sample. We expected that the shortened Chinese version of the MDRT would have satisfactory psychometric properties,

including high internal consistency reliability and satisfactory model fit of the latent factor model in the Chinese sample.

Methods

Participants The Chinese sample included 493[1] community adult participants (56.4% females, mean age $M = 27.74$, $SD = 11$). Participants were recruited via social networks, primarily over the WeChat social platform and in online social groups in April 2020. Most of the participants (63.9%) were either currently completing or had completed a bachelor degree, and a further 26.3% were either completing or had completed a postgraduate degree. The majority of participants identified as Han ethnicity (96.2%). About 66.1% of the participants were residents of Guangdong Province.

The US data for cross-cultural comparison was from Shou and Olney (2021). The sample contained 493 participants (48.3% females, mean age $M = 40.76$, $SD = 15.01$). Participants were recruited via the online survey platform Prolific, with 72.2% identifying as Caucasian and 14.1% as African American. A total of 57.2% of the participants had a tertiary education or higher. The US sample was recruited from the online crowdsourcing platform Prolific in September 2019 and September/October 2020. Further details about the US sample and the sampling procedure are reported in Shou and Olney (2021).

Materials

MDRT The MDRT scale that was developed in Study 1 was used in this study.

DOSPERT (Blais & Weber, 2006; Chinese version). The DOSPERT scale measures one's attitude toward risk in a range of situations across five domains: ethical, financial, health/safety, social and recreational. The scale consists of 30 items, with six items per domain subscale. Participants rated the likelihood that they would engage in each behavior on a 7-point scale from 1 = *extremely unlikely* to 7 = *extremely likely*.

---

[1] The sample size of 493 for the Chinese sample is sufficient for detecting the smallest hypothesized effect reported in the previous study (Shou & Olney, 2021; the relationship between alcohol consumption and health risk tolerance, $r = 0.16$) with a power of 0.8 and type 1 error of 0.05. The combined sample size ($N = 986$) is sufficient for performing network analysis (Golino & Epskamp, 2017; Shou & Olney, 2021)

Brief Sensation Seeking Scale (BSSS: Hoyle et al., 2002). Sensation seeking is an important construct that measures thrill seeking and impulsivity. The BSSS is a shortened version of the original sensation seeking scale and consists of eight items that assess one's tendency to engage in dangerous and thrilling activities rated on a 5-point scale from 1 = *strongly disagree* to 5 = *strongly agree*. The internal consistency reliability of the BSSS in the current study was satisfactory (alpha = 0.81, ωomega = 0.81). The Chinese version of the BSSS included items adapted from Chinese translations by Tseng (2010) and Lin (2012).

Financial Risk Tolerance Scale (FRTS: Grable & Lytton, 1999). The FRTS consists of 13 items that assess financial risks and investment preferences. The 13 items of the FRTS have a mixture of response categories and cover a range of financial investment and general financial risk attitude questions. The Chinese version of the FRTS used in this study was validated by Wang (2017).

TriPM Boldness Scale (Patrick, 2010). The TriPM is a scale based on the triarchical model of psychopathy. The boldness subscale assesses characteristics such as thrill-seeking, social adaptability and dominance, and fearlessness, which constitute the boldness trait. The scale consists of 19 items rated on a 4-point scale from 1 = *false* to 4 = *true*, with higher mean scores indicating a greater level of the boldness trait. The Chinese version of the TriPM has been validated in multiple Chinese samples and demonstrated satisfactory validity and reliability (Shou et al., 2016, 2017).

Brief Version of Fear of Negative Evaluation (BFNE: Leary, 1983). The BFNE is a 12-item scale that measures one's attitudes toward social and interpersonal negative evaluations. Aversion to negative evaluation is a key characteristic of social anxiety and social risk avoidance. The items were rated on a 5-point scale from 1 = *Not at all characteristic of me* to 5 = *Extremely characteristic of me*. A higher score on the BFNE indicated a greater fear of negative evaluation in social interactions and the average score of the BFNE was used in the analysis in this study. The Chinese version of the BFNE (Wang et al., 1999) was used in this study.

Health Behaviors We assessed the following five common health-related behaviors: smoking, drinking, regular health checks, physical exercise, and diet, as reported in Shou and Olney (2021). Smoking was assessed by participant smoking status (0 = *never smoked*, 1 = *former smoker* and 2 = *current smoker*) and frequency of smoking for current smokers (from 1 = *1 day or less a month*, 2 = *2-4 days a month*, 3 = *2-3 days a week*, 4 = *4-6 days a week*, and 5 = *Everyday*). Drinking behavior was measured by the total score of (1) frequency of drinking (0 = *never*, 1 = *1 day a month or less*, 2 = *2-4 days a month*, 3 = *2-3 days a week*, and 4 = *4 days or more a week*), (2) the amount consumed on a typical occasion (from 1 = *1 or 2 standard drinks* to 5 = *10 or more standard drinks*), and (3) the frequency of binge drinking (6 or more standard drinks on one single occasion; rated from 0 [*never*] to 4 [*4 days or more a week*]). Participants also indicated how often (from 1 [*never*] to 5 [*always*]) they engage in regular physical/health checks, exercise for at least 30 minutes a day 3 times a week, and eating at least 5 servings of fruit and vegetables per day.

Procedure The survey was programmed and conducted on the Chinese online survey platform wjx.cn, and the link to the survey was distributed via social networks. Participants who accessed the survey and consented to participate in the study completed the demographics information, DOSPERT, MDRT, FRTS, Boldness, BFNE, BSSS, and health behaviors questionnaire, in that order. The order of items within each scale was randomized. The median time taken to complete the survey was 15 minutes and participants received CNY5 (approximately AU$1.25) via WeChat pay upon completion of the survey.

Data Analysis

We first tested the construct validity of the version of MDRT (36 items) that was developed for the English-speaking samples (Shou & Olney, 2021) in the Chinese sample to investigate the necessity of establishing a joint version from the original item pool. Confirmatory factor analysis (CFA) with weighted least square estimation with adjusted mean and variance was used to test this.

Next, we applied Exploratory Graph Analysis (EGA) with graphical least absolute shrinkage and selection operator (LASSO) regularization on the combined Chinese and US sample to explore the joint clustering pattern among items (Golino & Epskamp, 2017; also see Shou & Olney, 2021 for more details on the application of EGA for scale construction). The EGA is advantageous over traditional exploratory factor analysis (EFA) in terms of accurately identifying dimensions when dimensions are correlated (Golino & Epskamp, 2017). We performed EGA using the 'EGAnet' package (Golino & Christensen, 2021) and the estimation was based on Spearman's correlations and a multi-level modularity optimization algorithm. The stability of the EGA classifications was tested using bootstrap simulations. Items that demonstrated a clear and stable clustering pattern were retained. Next, the network comparison test (NCT; van Borkulo et al., 2017) was carried out to test the equivalence of the network structure of the selected items across the two samples. We tested the equivalence of the overall network structure in terms of whether an edge's (i.e., links among

items) presence was identical between two networks (i.e., the same links among items were present in the network for both samples), strength of individual edges, and global strength estimates (i.e., the overall strength of the links among items in the network; Fried et al., 2018).

The network analysis approach (EGA and NCT) allows us to test the general dimensionality and clustering of the items and the equivalence of inter-item connections. The test of measurement invariance of the MDRT-EC as a measure of risk tolerance for the six domains also requires an understanding of how the items within each cluster represent their underlying latent factors across the two samples. Thus, we carried out multi-group confirmatory factor analysis (MG-CFA) to examine the extent to which the aggregated scores of items had an equivalent representation of domain-specific risk tolerance across the two samples.

We followed the most recent guideline by Svetina et al. (2020) based on Wu and Estabrook (2016)'s approach to model identification and measurement invariance testing for categorical indicators. There are four main steps of invariance testing. First, a baseline configural invariance model is built with an assumption that the factor-item combination is the same across models while all parameters are freely estimated. Next, a second model is built by constraining thresholds to be equal across groups, and a third model subsequently constrains the factor loading.[2] Metric invariance is achieved when the third model does not have significant change in model fit compared to the second model. Metric invariance is the prerequisite for comparing the MDRT-EC's correlations with external variables across groups. Finally, a fourth model is built by further constraining the intercepts of items. Strong, or scalar, invariance is achieved when the fourth model does not have significant change in model fit compared to the third model. Scalar invariance is the prerequisite for comparing the factor/subscale means across groups.

The MG-CFA models (for the six-correlated-factor structure) were estimated using weighted least square estimation with adjusted mean and variance. We focused on model fit indices, including scale shifted Comparative Fit Index (CFI), Tucker Lewis Index (TLI), and RMSEA. An insignificant change in the model fit between a more constrained and a less constrained model suggests satisfaction of the level of invariance added in the more constrained model. Changes in model fit smaller than 0.01 for CFI and TLI, and smaller than 0.005 for RMSEA, are considered insignificant (Chen, 2007; Maydeu-Olivares et al., 2018).

When metric invariance is satisfied, the correlations between MDRT-EC scales and criterion scales of the Chinese and US samples were then comparable. These correlations, which indicate the strength of association between MDRT-EC scales and criterion variables, would be first compared using Fisher's z transformation (Steiger's test). A significant test result suggests a significant difference in the strength of association between the two samples. When scalar invariance is satisfied, the mean scores of the MDRT-EC scales between Chinese and US samples were then compared. As age and gender differed between the two samples, we conducted linear regression analyses to control for the effect of age and gender. Each MDRT-EC domain scale was treated as the dependent variable and culture, age and gender were included as the independent variables.

Finally, we applied multiple regression to the investigation of the unique associations between MDRT-EC scales and each of the lifestyle health behaviors, as well as cultural differences. For each of the lifestyle behaviors, a full model was run including the six MDRT-EC domain scales, age and culture, as well as the interaction between MDRT-EC scales and age/gender/culture. Significant interactions were retained based on the stepwise process using BIC values. Multinomial regression was used for smoking status (non-smoker, past smoker and current smoker), linear regression was used for alcohol consumption, and ordinal regression was used for health check, exercise, and diet (measured on 5-point ordinal scales). Across all tests for which we did not have specific hypotheses (for Fisher'z tests, mean difference tests, regression models), p values were adjusted using the Benjamini-Hochberg procedure to control for Type 1 errors.

All analyses were carried out using the R (v3.6.2) program (R Core Team, 2016). EGA was performed using the 'EGAnet' package (Golino & Christensen, 2020) and the network comparison test was performed using the 'Network-ComparisonTest' package (van Borkulo et al., 2017). CFA was estimated using the 'lavaan' package (Rosseel, 2012) and the measurement invariance test was performed using the 'semTools' package (Jorgensen et al., 2021). Multinomial and ordinal regression models were estimated using the 'nnet' (Ripley & Venables, 2021) and 'ordinal' (Christensen, 2019) packages, respectively.

Results

CFA A six-factor model based on the items selected for the English samples did not achieve satisfactory model fit in the Chinese sample, scale shifted CFI = 0.89, TLI = 0.88, RMSEA = 0.073, 90% CI = [0.069, 0.076].

EGA An initial EGA on the 52 MDRT items revealed several cross-loading items that were similar to the results reported

---

[2] This is different from the traditional order of testing measurement invariance, in which factor loading is constrained, followed by threshold/intercepts. However the traditional approach could have model identification issues for categorical data as a factor loading invariant model might not be identified without constraining the threshold.

Table 1 Results of measurement invariance test

	χ^2	df	p	CFI	TLI	RMSEA
M1: Configural	4624	1690	<.001	0.920	0.914	0.059
M2: Equal Thresholds	4920	1858	<.001	0.916	0.919	0.058
M3: M2+ Equal Loadings	4878	1895	<.001	0.919	0.922	0.057
M4: M3 + Equal Intercepts	5398	1932	<.001	0.905	0.912	0.060
M5: M3 + Partial Equal Intercepts	5241	1930	<.001	0.91	0.915	0.059
		Δdf		ΔCFI	ΔTLI	ΔRMSEA
M2 vs M1		168		-0.003	0.004	-0.002
M3 vs M2		37		0.002	0.004	-0.001
M5 vs M4		37		0.014	0.010	0.003
M5 vs M3		35		-0.009	-0.007	0.002

The ratings of 7 and 6 were merged for four items (items 1, 4, 7, and 9) due to the overall small number of ratings of 7 and/or one group did not have responses that rated at 7

in Shou and Olney (2021). We removed cross-loading items and re-ran the EGA iteratively. A total of nine crossloading items were removed. Six of the nine items (Items 8, 17, 19, 26, 29 and 35) were also identified as cross-loading in Shou and Olney (2021). Two items (Items 12 and 18) were removed as a result of cross-loading in the joint EGA (items loaded on a cluster that was different from the initial domain allocation) and the last item (item 41) was removed due to not having stable clustering in bootstrapping stability analysis (< 90% replicability).

Bootstrapping simulations with 1000 samples for testing the network stability of the remaining 43 items suggested six clusters were identified in 88% of the bootstrapped samples. All of the items had relatively stable cluster allocations (same cluster allocation for 95% or more of bootstrapped samples). The NCT based on 5000 permutations indicated that the network structure ($p = .096$) and global network strength ($p = .456$) did not differ significantly between the Chinese and US samples. Only nine of the 903 (< 1%) edges had significant differences in their strength between the two samples. Overall, the 43 items had a similar network structure for the Chinese and US samples.

MG-CFA MG-CFA was carried out on the 43 items using a six-correlated factor model. Table 1 shows the results. The model fit of a configural invariance model was acceptable (scale shifted CFI = 0.920, TLI = 0.914, and RMSEA = 0.059, 90%CI = [0.057, 0.061]), indicating the item-latent factor combination among the 43 items and the 6 correlated latent factors were similar across the Chinese and US samples. The metric invariance model that constrained factor loading and thresholds across the two groups did not have a substantial change in model fit compared to a model that only constrains thresholds (ΔCFI and ΔTLI ≤ 0.005), indicating that the item-factor associations were generally equivalent across the two groups. However, a scalar invariance model that further constrained the item intercepts across

groups had a slight decrease in the indices of CFI and TLI (not in RMSEA) for change in model fit compared to the metric invariance model (ΔCFI = 0.012, ΔTLI = 0.011, and ΔRMSEA = 0.004). An inspection of the modification indices suggested that items 33 (skiing) and 51 (admitting different views) had substantially different item intercepts between the samples. Given the same level of the latent trait, observed scores for item 51 were significantly lower in the Chinese sample than in the US sample, while observed scores for item 33 were significantly higher in the Chinese sample than in the US sample. A partial scalar invariance model that relaxed the intercepts of items 33 and 51 did not have a substantial change in model fit compared to the metric invariant model (ΔCFI and ΔTLI ≤ 0.01, ΔRMSEA = 0.002). The results of the measurement invariance tests suggest partial strong invariance (equal factor loading and threshold, and equal intercepts except for two items) of the joint scale across the Chinese and US samples. We named the 43-item inventory as MDRT-EC.

Reliability Table 2 displays the descriptive statistics of the MDRT-EC in the Chinese sample and the statistics for the US sample can be found in supplementary Table S3. All six domain scales demonstrated satisfactory internal consistency (alpha/omega >=0.78; AIC values in the 0.2-0.5 range) in both samples. The discriminant validities of the MDRT-EC domain scales are supported by the substantially higher item-to-total correlation coefficients than the non-target item-to-total correlation coefficients in both samples.

Criterion Validity Invariance Given that measurement invariance of the MDRT-EC between US and Chinese samples was achieved, we examined cultural differences in the associations between the MDRT-EC domain scales and covariate scales. Table 3 displays the correlations between the six domain scales and other covariates for the Chinese sample, as well as the Fisher's z-tests for the differences in

Table 2 Descriptive statistics of variables for the Chinese sample

Variable	M	SD	median	Skew	alpha	omega	AIC	CITC	NTITC
MDRT									
Ethical	2.24	0.96	2.12	0.84	0.88	0.88	0.47	0.73	0.24
Financial	2.48	1.12	2.33	0.75	0.84	0.85	0.47	0.75	0.23
M/H	2.86	0.93	2.88	0.15	0.78	0.78	0.30	0.63	0.29
Social	3.67	1.00	3.62	0.05	0.82	0.82	0.36	0.66	0.28
RS	3.23	1.26	3.17	0.22	0.81	0.81	0.42	0.72	0.26
Recreational	4.96	1.01	5.00	-0.48	0.83	0.84	0.42	0.71	0.21
DOSPERT									
Ethical	2.33	0.91	2.17	0.66	0.63	0.67	0.22	0.61	0.15
Financial	3.15	1.12	3.00	0.57	0.78	0.80	0.37	0.69	0.19
Health	3.11	1.07	3.17	0.10	0.62	0.62	0.21	0.59	0.21
Social	4.78	0.92	4.83	-0.87	0.62	0.64	0.21	0.59	0.16
Recreational	3.15	1.26	3.17	0.16	0.76	0.77	0.35	0.67	0.20
Bold	2.45	0.38	2.47	0.16	0.86	0.86			
FNE	3.31	0.75	3.33	-0.31	0.92	0.92			
FRTS	24.17	3.84	24.00	0.55					
BSSS	2.70	0.75	2.75	0.17	0.79	0.79			
Health Check	2.55	1.17	2.00	0.84					
Exercise	2.89	1.16	3.00	0.33					
Diet	3.64	0.93	4.00	-0.71					
Drink Total	1.67	1.96	2.00	1.06					
Smoke Total	0.43	1.37	0.00	4.07					

MDRT = Multi-Domain Risk Tolerance scale. M/H = medical/health domain. RS = recreational-safety domain. Bold = Boldness scale. FNE = Fear of negative evaluation scale. BSSS = Brief sensation seeking scale. FRTS = Financial Risk Tolerance Scale. alpha = Cronbach's alpha. omega = omega total. AIC = average inter-item correlation. CITC = corrected item-to-total correlations. NTITC = non-target item-to-total correlations.

the correlations between the Chinese and the US samples. As expected, the MDRT-EC domain scales had significantly positive correlations with corresponding DOSPERT domain scales, and the correlations with the corresponding DOSPERT domain scales were stronger than correlations with non-corresponding domain scales. In addition, the MDRT-EC financial domain significantly and positively correlated with the FRTS ($r = 0.33$), the recreational-safety domain had a significant and positive correlation with sensation seeking ($r = 0.55$), the MDRT-EC social domain had a significant and positive correlation with boldness ($r = 0.31$) and had a significant and negative correlation with fear of negative evaluation ($r = -0.18$). Most target associations (i.e., in bold) were not significantly different between the US and Chinese samples, including the direction and magnitude of the associations. This provides further evidence for the measurement invariance of the MDRT-EC.

Cultural Differences in Risk Attitudes and Predictions of Lifestyle Behaviors Table 4 shows the test results comparing the mean scores of the MDRT-EC and DOSPERT scales (excluding items 33 and 51) between the Chinese and US

samples. After controlling for age and gender, the Chinese sample had significantly higher scores on the health risk tolerance and recreational risk tolerance domain scales. There were no significant differences in the mean scores for the other four domains.

The results of the associations between MDRT-EC scales and lifestyle behaviors for the Chinese sample and cultural differences are displayed in Table 5. Alcohol consumption was positively predicted by ethical ($b = 0.38$) and recreational-safety ($b = 0.28$) risk tolerance. Regular health checks were negatively associated with medical/health risk tolerance ($b = -0.43$), while positively associated with social risk tolerance ($b = 0.32$). The link between social risk tolerance and health checks was also significantly different between the Chinese and US sample ($b = -0.01$ for the US sample). Finally, physical exercise was significantly and positively predicted by social risk tolerance ($b = 0.39$), while it was significantly and negatively predicted by ethical ($b = -0.23$) and recreational ($b = -0.23$) risk tolerance. On the other hand, risk tolerance did not significantly predict dietary behavior or smoking status. There were no significant

Table 3 Correlations between MDRT domain scales and criterion scales for the Chinese sample and group differences

Variable	Ethical	Financial	M/H	Social	RS	Recreational
MDRT						
Financial	0.46**					
M/H	0.51**	0.44**				
Social	0.30**	0.36**	0.36**			
RS	0.34**	0.39**	0.41**	0.45**		
Recreational	0.10*	0.15*	0.26**	0.48**	0.34**	
DOSPERT						
Ethical	**0.64****	0.33**	0.34**	0.19**	0.22**	0.06
	(0.04)	(0.07)	(0.10)	(0.02)	(0.14)	(0.05)
Financial	0.33**	**0.74****	0.28**	0.32**	0.32**	0.16**
	(-0.01)	(-0.04)	(-0.06)	(-0.05)	(0.02)	(-0.1)
Health	0.44**	0.32**	**0.41****	0.27**	**0.37****	0.13*
	(-0.04)	(0.02)	(0.01)	(0.02)	(0.12)	(0.08)
Social	0.04	0.21**	0.15*	**0.38****	0.32**	0.38**
	(-0.03)	(-0.11)	(-0.07)	(0.05)	(-0.12)	(-0.07)
Recreational	0.29**	0.37**	0.3**	0.33**	**0.69****	**0.15***
	(0.03)	(-0.03)	(0.01)	(-0.02)	(-0.02)	(0.03)
FRTS	0.10*	**0.38****	0.15*	0.26**	0.28**	0.15*
	(0.18)	(0.05)	(0.04)	(0.04)	(0.04)	(-0.08)
BSSS	0.29**	0.28**	0.18**	0.35**	**0.55****	0.17**
	(0.07)	(0.07)	(0.12)	(-0.02)	(0.02)	(0.07)
Bold	-0.02	0.21**	0.04	**0.31****	0.27**	0.13*
	(0.12)	(-0.05)	(-0.06)	(0.1)	(-0.08)	(-0.1)
FNE	0.01	0.17**	-0.02	**-0.18****	-0.16*	0.03
	(-0.01)	(0.16)	(0.1)	(-0.07)	(0.21)	(0.05)

*$p < .05$, **$p < .001$. MDRT = Multi-domain risk tolerance scale. M/H = medical/health domain. RS = recreational-safety domain. Bold = Boldness scale. FRTS = Financial risk tolerance scale. BSSS = Brief sensation seeking scale. FNE = Fear of negative evaluation. The coefficients in bold are hypothesized correlations. The values in parentheses are the differences between the correlation in the US sample and the one in the Chinese sample ($r_{US} - r_{Chinese}$). A positive value indicates the correlation was more positive for the US sample than the Chinese, while a negative value indicates the correlation was more negative for the US sample than the Chinese sample. Fisher z tests' p values were adjusted based on 54 pairs of tests.

Table 4 Mean comparison between Chinese and US samples on the MDRT domain scales

Subscale	Mean difference (US - Chinese)	t	d	b
Ethical	-0.18	-2.84*	-0.18	-0.04
Financial	0.01	0.13	0.01	-0.02
Medical/Health	-0.36	-6.04**	-0.39	-0.34**
Social	0.03	0.52	0.03	0.06
Recreational-Safety	-0.29	-3.5**	-0.22	-0.06
Recreational	-0.5	-7.63**	-0.49	-0.32**

*$p <.05$, **$p< .001$. The b coefficients (unstandardized) indicate the effect of culture on the MDRT domain scale with age being controlled for in a regression model. All p values were adjusted based on 6 tests.

differences in the links between risk tolerance and physical exercise, diet and smoking between the two samples.

Short Version of Chinese MDRT Using the 43-item MDRT-EC, we selected items that performed best for the Chinese sample to develop a short version of the Chinese MDRT for future research in Chinese-speaking populations. Supplementary Table S4 shows the final version of the Chinese MDRT (named MDRT-C). The six-correlated factor of the MDRT-C had satisfactory fit and the factor loading of all items were above 0.5 (see supplementary Table S5). The internal consistency, discriminant validity and criterion validity of the MDRT-C are comparable to that of the MDRT-EC for the Chinese sample (Table S6 and S7).

Discussion

The results suggested that certain items selected for the English-speaking sample might not function as well in the Chinese sample. Thus, we developed a version of the MDRT, which we have termed MDRT-EC, that could perform similarly for both Chinese and US samples. We addressed

Table 5 Regression results predicting health behaviors from MDRT domain scales and culture

Variable	Drinking	Diet	Exercise	Health check	Past vs Non-smoker	Current vs Non Smoker
Ethical	0.38**	-0.09	-0.23*	-0.1	0.17	0.06
Financial	0.13	0.16	0.14	0.16	0.05	0.20
M/H	0.05	-0.13	-0.06	-0.43**	0.14	0.12
Social	-0.01	0.05	0.39**	0.32**	0.05	0.08
RS	0.28**	-0.04	0.04	-0.01	0.15	0.12
Recreational	-0.03	0.11	-0.23*	0.11	0.02	-0.30
Age	0.02*	0.01	0.01*	0.05**	0.04**	0.03*
Gender	-0.35**	0.16	-0.07	0.20*	-0.24*	-0.28*
Culture (US vs. Chinese)	1.13**	-2.15**	-0.54**	-0.43*	0.70**	1.71**
M/H x Age				-0.02*		
Recreational x Age				0.02**		
RS x Gender					-0.30**	-0.26*
Social x Culture				-0.33*		

*$p < .05$, **$p < .001$. All p values were adjusted based on 5 tests. M/H = medical/health domain. RS = recreational-safety domain. The coefficients for diet, exercises and health check were ordinal regression coefficients. The coefficients for alcohol consumption were Poisson regression coefficients. The coefficients for smoking status are multinomial regression coefficients with non-smokers treated as the base group. Age was centred and all MDRT scales were standardized. For Culture, the Chinese sample was coded as 0 and the US sample was coded as 1. For Gender, females were coded as 1 and males as -1

different levels of measurement invariance, including item dimensionality and clustering, inter-item connectedness, latent factor structure and criterion validity. The establishment of metric invariance indicates that the items have similar predictive power to the latent factors across the two groups. The partial strong invariance between the two samples indicates that the same level of the latent trait would predict the same observed scores for participants from both groups (except for two items). Further construct invariance was demonstrated by the two samples having similar convergence between MDRT-EC scales and conceptually relevant scales.

We then compared risk attitudes across domains between the two groups with the items that had strong measurement invariance. Chinese participants were substantially more tolerant of risk in the recreational and health domains than the US participants when gender and age effects had been accounted for. As suggested by the cushion hypothesis (Hsee & Weber, 1999), risks can be shared by group members of the individuals' social support network. For instance, social norms make it more likely that financial loss can be overcome by borrowing money from others in the Chinese cultural context. Although not explicitly stated within the cushion hypothesis, it follows that certain health risks may also be mitigated by one's social network, such as being taken care of by one's family members. As such, Chinese participants may show a higher tolerance of risk, as there is a stronger norm of maintaining social support networks. At the same time, Chinese participants were less tolerant of

social risks as social connections are the buffer that allows them to deal with risk in other domains and they may thus perceive greater risk in the social domain because of this.

In terms of the associations between lifestyle behaviors and MDRT-EC domain scales, most results were similar to the ones reported in Shou and Olney (2021), as there were no significant differences in the associations between risk tolerance and lifestyle behaviors in most cases, with one exception. Social risk tolerance had a positive association with health checks in the Chinese sample, while this link was not significant for the US sample. One possible explanation for this is the cultural differences in doctor-patient interactions. In the US system, patients usually have private consultations with a family doctor. By contrast, instead of having regular family doctors, most patients in the Chinese cultural setting visit public hospitals for health checks. The doctors can be complete strangers and the consultation can be less private than that facilitated by the US system. Thus, tolerance of social interactions and settings in general may be conducive to health checks with unfamiliar people and being exposed to more public spaces.

General Discussion

Many previous studies on cultural differences in risk attitudes did not account for cross-cultural measurement issues. Few studies have directly investigated cultural differences in risk attitudes between Euro-Americans and East Asians

across multiple domains with a cross-culturally validated measure. We established a joint version of the Multi-Domain Risk Tolerance scale for both Chinese and US samples to enable more appropriate cross-cultural comparison. The comparison of the mean scores of the scales and the prediction of lifestyle behaviors were performed only when measurement invariance was achieved and when the measure demonstrated satisfactory psychometric properties in both groups. The results demonstrated that Chinese participants scored significantly higher on recreational and health risk tolerance than the US participants. In addition, the two groups had significant differences in the associations between health risk tolerance and alcohol consumption, and between social risk tolerance and health checks.

There are several implications of the findings in the current study. First, the current study highlights the importance of measurement invariance when the risk attitude scale is applied to understand cultural differences in risk attitudes. At face value, the current findings on cultural differences in risk tolerance may seem different to some of the conclusions found in previous studies (e.g., Park et al., 2015; Weber & Hsee, 1998; Terpstra-Tong & Terpstra, 2013). However, the previous studies were based on measures of which the mean scores might not be directly comparable (e.g., Park et al., 2015), based on choice behaviors (Weber & Hsee, 1998), or involved behavioral tendency and required prior knowledge (Terpstra-Tong & Terpstra, 2013). From these previous studies alone it is difficult to infer cultural differences in the affective component of risk attitudes, and therefore the findings are not directly comparable with the findings in the current paper.

While a number of studies demonstrate that there are cultural differences in perceived risks across different domains, cultural influences on risk attitudes – beyond the effects it has on perceived risk – are yet to be fully explored. A number of questions remain unanswered. For example, how different cultural groups tolerate risk across domains when the perceived risks in a situation are similar? Are relationships among risk attitudes across domains the same across different groups, and, if not, what are the reasons? Are consequences and causes of domain-specific risk attitudes the same across cultural groups? The current results, by using a culturally invariant tool, suggest that the extent to which one tolerates risk in a specific domain may not only relate to the degree of perceived risk, but also to the perceived ability to handle the loss.

Second, the joint MDRT and the subsequently constructed Chinese short version of the scale show good psychometric properties, including latent factor validity and internal consistency for Chinese participants. This suggests that items are less influenced by familiarity, prior experience, and personal relevance when being applied to the Chinese cultural context. This finding also implies that specifying the nature

and domain of the risk in the item can reduce the impact of individual differences that are irrelevant to risk attitudes when items are applied to any new group.

Third, the findings on the associations between risk tolerance and lifestyle behaviors highlight the complexity of real-life behaviors. While there is no doubt that all five lifestyle behaviors, whether engaging in them or not engaging in them, in the current study involve health risks, their significant predictors can be risk attitudes towards other life domains, such as the social, ethical and recreational domains. The consideration of the other domains' risk could depend on the presence of the risk in a certain social context, such as legal regulation, public awareness of health risks, and the nature of the public health and medical system.

One limitation of the current study is that all measures were self-report. The accuracy of some variables, such as lifestyle behaviors, are therefore subject to measurement biases such as the social desirability bias. Future studies could include more specific measures or objective indicators of real-life behaviors in order to more comprehensively assess risk tolerance across domains and their practical implications. Second, most of the participants in Study 2 were from urban areas (e.g., Guangzhou) and many participants were young and highly educated. This could limit the generalizability of the current results to wider populations with diverse education and age distributions. There is a need for more replication studies, especially involving samples that are more representative or have more closely matching demographic characteristics between two comparison samples. Third, the current study accounted for the applicability of the items to the Chinese and US cultural context by selecting items from a 52-item pool, which was constructed in light of Euro-American and East Asian views on domain-specific risks. However, the applicability of the items for appropriate assessment of other important cultural contexts, such as Indian, Arabic and African, requires further investigation. Finally, the current study focused on validity from internal structure (e.g., reliability and factor structure) and convergence with related constructs. Future studies should consider measuring a wider range of outcome variables across domains and assess both the predictive and incremental validity of the MDRT.

In conclusion, the current study demonstrated that the MDRT could be a promising tool to measure risk tolerance across cultures, and to understand cultural differences in risk attitudes and constructs related to risk attitudes. The domain-specific framework of risk attitudes is also useful to understand how risk considerations in health and lifestyle behaviors can differ across different cultural and ethnic groups. Understanding cultural differences in risk attitudes would also benefit a range of applied areas. For risk communication, designing group-specific messaging for culturally-diverse populations that accounts for different levels

of risk tolerance would be one way to apply the findings of this paper and create greater efficacy in communicating risk. Similarly, in the field of health promotion, designing campaigns or measures tailored to specific cultural groups may be a more effective way to reach equitable health goals across these populations. Finally, in the broader area of cross-cultural/national politics, acknowledging inherent differences in risk tolerance between participating parties may help to alleviate tension and conflict that arises between groups of people that manifest different views on important issues. Future research should further investigate the causes and consequences of cultural differences in risk tolerance.

Declarations

References

Blais, A.-R., & Weber, E. U. (2006). A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making, 1*(1), 33–47.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, H. Y., & Tao, J. (2013). Risk perception and risk-taking attitude: A comparison between Hong Kong and mainland Chinese undergraduate students. *The Asia-Pacific Education Researcher, 22*(4), 497–506. https://doi.org/10.1007/s40299-012-0048-7

Cheung, H. Y., Wu, J., & Tao, J. (2016). Predicting domain-specific risk-taking attitudes of mainland China university students: A hyper core self-evaluation approach. *Journal of Risk Research, 19*(1), 79–100. https://doi.org/10.1080/13669877.2014.948903

Christensen, R. H. B. (2019). ordinal: Regression Models for Ordinal Data. Retrieved from https://cran.r-project.org/web/packages/ordinal/ordinal.pdf

Du, X., Li, J., & Du, X. (2014). Testing risk-taking behavior in Chinese undergraduate students. *PLoS ONE, 9*(5), e97989.

Figner, B., & Weber, E. U. (2011). Who takes risks when and why? Determinants of risk taking. *Current Directions in Psychological Science, 20*(4), 211–216. https://doi.org/10.1177/0963721411415790

Fried, E. I., Eidhof, M. B., Palic, S., Costantini, G., Huisman-van Dijk, H. M., Bockting, C. L., Engelhart, I., Armour, C., Nielsen, A. B. S., & Karstoft, K. I. (2018). Replicability and generalizability of posttraumatic stress disorder (PTSD) networks: A cross-cultural multisite study of PTSD symptoms in four trauma patient samples. *Clinical Psychological Science, 6*(3), 335–351. https://doi.org/10.1177/2167702617745092

Golino, H., & Christensen, A. P. (2021). EGAnet: Exploratory Graph Analysis– a framework for estimating the number of dimensions in multivariate data using network psychometrics. https://CRAN.R-project.org/package=EGAnet

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PloS one, 12*(6), e0174035.

Grable, J., & Lytton, R. H. (1999). Financial risk tolerance revisited: the development of a risk assessment instrument. *Financial Services Review, 8*(3), 163–181. https://doi.org/10.1016/S1057-0810(99)00041-4

Hoyle, R. H., Stephenson, M. T., Palmgreen, P., Lorch, E. P., & Donohew, R. L. (2002). Reliability and validity of a brief measure of sensation seeking. *Personality and Individual Differences, 32*(3), 401–414. https://doi.org/10.1016/S0191-8869(01)00032-0

Hsee, C. K., & Weber, E. U. (1999). Cross-national differences in risk preference and lay predictions. *Journal of Behavioral Decision Making, 12*(2), 165–179. https://doi.org/10.1002/(SICI)1099-0771(199906)12:2%3c165::AID-BDM316%3e3.0.CO;2-N

Hu, X., & Xie, X. (2012). Validation of the Domain-Specific Risk-Taking Scale in Chinese college students. *Judgment and Decision Making, 7*(2), 181–188.

Jorgensen, T. D., Pornprasertmanit, Schoemann, A., & Rosseel, Y. (2021). semTools: Useful tools for structural equation modeling. Retrieved from https://cran.rproject.org/web/packages/semTools

Kim, D. Y., & Park, J. (2010). Cultural differences in risk: The group facilitation effect. *Judgment and Decision Making, 5*(5), 380–390.

Leary, M. R. (1983). A brief version of the Fear of Negative Evaluation Scale. *Personality and Social Psychology Bulletin, 9*(3), 371–375. https://doi.org/10.1177/0146167283093007

Lin. (2012). Brief Measures of Sensation Seeking: Scale Development and Validation. *Journal of Health Promotion and Health*

Education, 37, 23–53. https://doi.org/10.7022/2fJHPHE.201206.0023

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(3), 389–402. https://doi.org/10.1080/10705511.2017.1389611

Park, J., Kim, D.-Y., & Zhang, C. (2015). Understanding cross-national differences in risk through a localized cultural perspective. *Cross-Cultural Research, 50*(1), 34–62. https://doi.org/10.1177/1069397115609560

Patrick, C. J. (2010). *Operationalizing the triarchic conceptualization of psychopathy: Preliminary description of brief scales for assessment of boldness, meanness, and disinhibition* [Unpublished test manual, Florida State University]. Retrieved from https://patrickcnslab.psy.fsu.edu/wiki/images/b/b2/TPMmanual.pdf

Pizam, A., Jeong, G. H., Reichel, A., van Boemmel, H., Lusson, J. M., Steynberg, L., State-Costache, O., Volo, S., Kroesbacher, C., Kucerova, J., & Montmany, N. (2004). The relationship between risk-taking, sensation-seeking, and the tourist behavior of young adults: A cross-cultural study. *Journal of Travel Research, 42*(3), 251–260. https://doi.org/10.1177/0047287503258837

R Core Team. (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

Reisinger, Y., & Mavondo, F. (2006). Cultural differences in travel risk perception. *Journal of Travel & Tourism Marketing, 20*(1), 13–31. https://doi.org/10.1300/J073v20n01_02

Ripley, B., & Venables, W. (2021). nnet: Feed-forward Neural Networks and Multinomial Log-Linear Models. Retrieved from https://cran.r-project.org/web/packages/nnet/nnet.pdf

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more Version 05–12 (BETA). *Journal of Statistical Software, 48*(2), 1–36.

Shou, Y., & Olney, J. (2020). Assessing a domain-specific risk-taking construct: A meta-analysis of reliability of the DOSPERT scale. *Judgment and Decision Making, 15*(1), 112–134.

Shou, Y., & Olney, J. (2021). Measuring risk tolerance across domains: Scale development and validation. *Journal of Personality Assessment*. https://doi.org/10.1080/00223891.2021.1966019 Online first publication.

Shou, Y., Sellbom, M., & Han, J. (2016). Development and validation of the Chinese Triarchic psychopathy measure. *Journal of Personality Disorders, 30*(4), 436–450. https://doi.org/10.1521/pedi.2016.30.4.436

Shou, Y., Sellbom, M., Xu, J., Chen, T., & Sui, A. (2017). Elaborating on the construct validity of Triarchic Psychopathy Measure in Chinese clinical and nonclinical samples. *Psychological Assessment, 29*(9), 1071–1081. https://doi.org/10.1037/pas0000398

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

Terpstra-Tong, J. L. Y., & Terpstra, R. H. (2013). Differences in financial risk tolerance: A study of the United States, China, Hong Kong and Macao. *Euro Asia Journal of Management, 42*(23), 43–65.

Tseng, C.H. (2010). *Does sex kill or heal? Influences of types of sexual appeals, product type and sensation seeking in advertising* [master's thesis, National Sun Yat-sen University, Taiwan]. eThesys. Retrieved from http://etd.lib.nsysu.edu.tw/ETD-db/ETD-search/view_etd?URN=etd-0310111-001145

van Borkulo C. D., van Bork R., Boschloo L., Kossakowski J., Tio P., Schoevers R. A., Borsboom D., & Waldorp L. J. (2017). *Comparing network structures on three aspects: A permutation test* [unpublished manuscript]. https://doi.org/10.13140/RG.2.2.29455.38569

Wang C.S. (2017). *Determinants of investment behavior: The roles of personality traits and risk attitude* [master's thesis, National Taiwan Normal University, Taiwan]. DSpace. Retrieved from http://rportal.lib.ntnu.edu.tw/handle/20.500.12235/94651

Wang, X., Wang, X. L., & Ma, H. (1999). Mental health assessment scale manual (updated version). *Chinese Journal of Mental Health, 13*, 318–320.

Wang, Y., Jing, Y., Zhang, Z., Lin, C., & Valadez, E. A. (2017). How dispositional social risk-seeking promotes trusting strangers: Evidence based on brain potentials and neural oscillations. *Journal of Experimental Psychology: General, 146*(8), 1150–1163. https://doi.org/10.1037/xge0000328

Weber, E. U., Blais, A. R., & Betz, N. E. (2002). A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making, 15*(4), 263–290. https://doi.org/10.1002/bdm.414

Weber, E. U., & Hsee, C. (1998). Cross-cultural differences in risk perception but cross-cultural similarities in attitudes towards perceived risk. *Management Science, 44*(9), 1205–1217.

Weber, E. U., Hsee, C. K., & Sokolowska, J. (1998). What folklore tells us about risk and risk taking: Cross-cultural comparisons of American, German, and Chinese proverbs. *Organizational Behavior and Human Decision Processes, 75*(2), 170–186.

Weber, E. U., & Johnson, E. J. (2009). Mindful judgment and decision making. *Annual Review of Psychology, 60*, 53–85. https://doi.org/10.1146/annurev.psych.60.110707.163633

Wichary, S., Pachur, T., & Li, M. (2015). Risk-taking tendencies in prisoners and nonprisoners: Does gender matter? *Journal of Behavioral Decision Making, 28*, 504–514. https://doi.org/10.1002/bdm.1866

Wu, J., & Cheung, H. Y. (2014). Confirmatory factor analysis of DOSPERT scale with Chinese university students. *Psychological Reports, 114*(1), 185–197. https://doi.org/10.2466/09.03.PR0.114k15w2

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