# Research in the Social Scientific Study of Religion

*Edited by*

Andrew Village and Ralph W. Hood Jr.


BRILL

LEIDEN | BOSTON

# Contents

# Chinese Psychology of Religion Measures: A Systematic Review and Best Practice Guidelines

*Kenneth T. Wang, Li Zhang and Yanmei Cao**

**Abstract**

The present systematic review identified scale development and psychometric evaluation articles in the field of Chinese psychology of religion (PR). Articles were searched in three databases, China National Knowledge Infrastructure (CNKI), PerioPath, and PsycINFO to locate PR studies published respectively in Mainland China, Taiwan, and internationally. In the 1,682 articles screened, 26 articles (22 unique measures) met the inclusion criteria. These 26 studies were then subjected to an evaluation of their scale development process and psychometric properties. A review of the studies indicated that the majority of measures showed evidence of adequate psychometric properties. However, there were also areas identified in which researchers could improve on, such as incorporating a more comprehensive review of culturally relevant literature and theory, providing more detailed descriptions of the item development process, construct Chinese indigenous scales rather than translating or adapting existing Western measures, and conducting more comprehensive evaluations of psychometric properties (e.g., conduct both exploratory and confirmatory factor analysis, examine construct validity). The systematic review also identified an increasing trend of new Chinese indigenous measures being developed, though the number is still low compared to translated Western measures. We encourage PR researchers to further develop culturally indigenous measures to establish a strong foundation for future research in this field. Guidelines on best practices for scale development were provided with an emphasis on cultural sensitivity.

**Keywords**

guideline – measures – psychology of religion – psychometric properties – review – scale

* *Author Note:* Kenneth T. Wang, Fuller Theological Seminary; Li Zhang and Yanmei Cao, Travis Research Institute, Fuller Graduate School of Psychology. Correspondence concerning this article should be addressed to: Dr. Kenneth T. Wang, Department of Clinical Psychology, Fuller Theological Seminary, 135 N. Oakland Ave, Pasadena, CA 91101, USA. Tel: 626-304-3712; Email: ktwang@fuller.edu.

The foundation of scientific research is to have accurate and appropriate assessment tools (Hill & Edwards, 2013). This has also been a key concern for the field of psychology of religious studies. Compared to Western psychology of religion, the field of psychology of religion is in its infancy in China (Dueck & Han, 2012). About two decades ago, Hill and Hood (1999) reviewed the measures used in the Western psychology of religion (PR) literature and identified 125 scales which they included in their edited book *Measures of Religiosity*. These measures were classified into 17 categories: religious beliefs and practices, religious attitudes, religious orientation, religious development, religious commitment and involvement, religious experience, religious/moral values or personal characteristics, multidimensional religiousness, religious coping and problem-solving, spirituality and mysticism, god concept, religious fundamentalism, death/afterlife, views of divine intervention/religious attribution, forgiveness, institutional religion, and related constructs. This resource book has allowed researchers ready access to psychological measures of religiosity to be used in research.

More recently, Hill and Edwards (2013) conducted another review of the measurements on psychology of religiousness and spirituality in the *APA Handbook of Psychology, Religion, and Spirituality*. In this updated review, they estimated that a total of 225 PR related measures were now available. They classified these measures into two broad categories—substantive measures and functional measures. Substantive measures focused on individuals' religious tendencies, beliefs, and behaviours. Under the substantive measure category, subtypes included general religiousness or spirituality, beliefs and religious preferences, religious or spiritual commitments, relational measures, religious or spiritual development, religious social participation, and private religious or spiritual practices. In contrast, functional measures focused on assessing the role religious characteristics and activities played in individuals' lives. Within this category, subtypes included religious motivation, religion as a source of meaning and values, religious support, spiritual experiences, religious coping, and religious struggle and strain. These categories provide a rough framework of the constructs and aspects involved in Western PR research. However, due to differences between Chinese and Western cultures, the relevance and importance of religious constructs could also differ. Therefore, cultural context is an important factor that should not be overlooked.

Caution should be taken when research on psychology is transported from one culture to another. Despite the benefit of having reviews on PR measures in 1999 and 2013 from the Western literature available as a potential reference/roadmap for Chinese PR research, Hill and Edwards (2013) pointed out that a main problem in the international psychology of religion field is related to cul-

tural biases. For example, the majority of religious and spiritual measures have a biased representation towards a Judeo-Christian context and a disproportion number focused on European American Protestants. More specifically, a Western measure on religious maturity (with European American Protestants) would probably not be appropriate or relevant when examining the religious maturity of Chinese Buddhists.

There are significant differences in the role of religion in cultures between China and the United States. First, polls have consistently shown that in the United States more than 70% of the population claim to have a religious belief and that the majority of the religious population is Protestant Christians (Gallup, 2015). In contrast, a recent poll showed that among the 57 countries investigated, China had the smallest portion (14%) that reported to be religious and the largest portion (47%) that reported to be atheists (WIN-Gallup, 2012). Despite the fact that religious people only make up a small portion of the Chinese population, the number of religious individuals in China is close to that in the United States due to the large population size of China. Second, observers have long argued that Protestant Christianity has fundamentally shaped American culture and beliefs since the first group of Puritans set foot on North America (Tocqueville, 1835/2003). In contrast, China consists of multiple cultural and ethnic groups, which leads to various types of religious groups and the absence of any single dominating religion. The Chinese government officially recognizes five religions: Buddhism, Catholicism, Daoism, Islam, and Protestantism. In addition, a large portion of Chinese believe in traditional Chinese folk beliefs and religions which are hard to map onto a believer/non-believer dichotomy (Wu, 2007). Due to the unique characteristics of religions in China, an indigenous approach toward the development of PR measures is necessary (Dueck & Han, 2012).

Many Chinese scholars have emphasized the importance of developing an indigenized Chinese psychology that takes into account the social, cultural, and intellectual context of the Chinese people (e.g., Yang, 1999). This indigenous approach contrasts with the psychological approach that is often employed to study Chinese populations which involves the utilization of constructs and measures developed in Western and highly individualistic contexts (Yang, 2006).

Uncritical use of Westernized psychological instruments could result in incompatibility problems, such as linguistic issues, inappropriate items, and construct under-representation. Linguistic issues are related to the different ways of expressing and wording scale items across languages. Despite the efforts expended in conducting an accurate translation through a community approach of translators and using back-translation to verify accuracy, the

reading difficulty, naturalness, and perceived meaning of the items in the translated form may differ from the original form (van de Vijver & Leung, 1997). Inappropriate items are the result of cultural differences due to values, customs, and environmental factors. For example, an item '*I feel comfortable using someone's first name soon after I meet them, even when they are much older than I am*' from the Singelis Self-Construal Scale (Singelis, 1994) that measured a respondent's level of independence was found to be inappropriate for respondents from East Asian cultures. To illustrate, in Chinese cultural settings it is extremely rare and impolite to address those much older by their first name. In other words, this item would be measuring 'social inappropriateness' rather than 'independent self-construal' with Chinese respondents. After realizing this issue, Singelis suggested changing the item to '*I can talk openly with a person who I meet for the first time, even when this person is much older than I*' when the scale is used for East Asian populations.

Construct underrepresentation could be another issue even after inappropriate items are adapted (Heppner, Wampold, Owen, Thompson, & Wang, 2015). Various constructs are weighed differently across cultures. For example, there is much emphasis on the concept of *yuan* and *ming* (different types of fate) in the Chinese culture compared to the Western. The idea of fate is commonly used among Chinese when making attributions. However, a Western attribution measure may emphasize only individual and environmental factors while neglecting the importance of fate. Thus, using the Western attribution measure to study Chinese populations may result in missing certain critical aspects of how Chinese people make attributions.

To facilitate the development of this emerging field in China, it would be particularly helpful for researchers to be aware of the PR measures available as well as the quality of these measures. A systematic review of the current PR measurement landscape of the Chinese research literature is a critical first step. The purpose of this content review is threefold. First, we will provide a general overview of the measures used to conduct PR research in China. We will present a list of the measures so that researchers can easily identify and locate relevant scales for their research. We will also make note of indigenous PR scales that were originally developed for and with Chinese populations. Second, we will evaluate current practices in the development and psychometric evaluation of Chinese PR measures by examining the procedures involved in the selected studies. Overall strengths and weakness will be summarized along with suggestions for future research. Third, we will address the key steps necessary for developing Chinese indigenous measures. The guideline of best practices for scale development will include how to: conduct a literature review, collect qualitative data to inform item development, write up strong scale items, finalize the scale through data analysis, and evaluate its psychometric properties.

## Method

### *Search Strategy and Process*

To identify scale development and psychometric evaluation studies published in Chinese and international journals, we conducted systematic searches in three databases. The goal of our study was to incorporate scale development articles with Chinese participants published not only in Mainland China, but also in Taiwan, and internationally. For papers published in Mainland China, we searched CNKI (China National Knowledge Infrastructure). For papers published in Taiwan, we searched PerioPath (Index to Taiwan Periodical Literature System). We also searched PsycINFO (American Psychological Association) for papers published internationally in English.

We used two different sets of search terms due to the language differences for the Chinese and English databases. Our goal was to find articles that were related to religion, and mainly focused on scale development or psychometric evaluation. For CNKI, we used the following terms in Simplified Chinese to search in the abstract: (AB = 宗教 <religion> OR 信仰 <faith> OR 灵性 <spirituality> OR 精神性 <spirituality> OR 民间信仰 <folk religion> OR 基督教<Christianity> OR 佛教<Buddhism> OR 道教<Daoism> OR 天主教<Catholicism> OR 伊斯兰教<Islamism> OR 穆斯林<Muslim> OR 新教 <Protestantism>) AND (AB = 量表 <scale> OR 问卷 <survey>) AND (AB = 信度 <reliability> OR 效度 <validity> OR 信效度 <psychometric properties> OR 因素分析 <factor analysis> OR 编制 <development> OR 建构 <construction> OR 修订 <adaptation> OR 验证 <validation> OR 检验 <examination> OR 自编 <originally developed> OR 翻译 <translated> OR 改编 <modified>). For PerioPath, we used the same set of search terms, but in Traditional Chinese. For PsycINFO, the following combination of search terms were used to search in article abstracts (scale OR inventor* OR question* OR measure*) AND (develop* OR valid* OR reliab* OR psychomet* OR initial*) AND (religio* OR spiritu* OR faith* OR christi* OR protesta* OR cathol* OR buddhi* OR musli* OR islam* OR taoi* OR daoi*) AND (Chines* OR Taiwan* OR Hong Kon* OR China*). The final search was conducted on 28 December 2015.

### *Inclusion/Exclusion Criteria*

To define the scope of our review study, we used the following inclusion/exclusion criteria:

1. Journal articles of empirical studies were included. Other types of publications (e.g., literature reviews, doctoral dissertations, and master's theses) were excluded.

2. Articles published with Chinese participants from Mainland China, Hong Kong, or Taiwan were included. Studies with other samples or only a proportion of which were Chinese were excluded.
3. Articles examining psychometric properties that included both reliability and validity were included. Articles with no psychometric property evaluation or those that only reported reliability statistics were excluded.
4. Scales that focused on religion and spirituality were included. Scales that focused on issues or settings unrelated to religion or spirituality, such as superstition, Chinese philosophy, industrial/organizational psychology, or classroom settings were excluded.
5. Scales that were developed with religion or spirituality as the main focus were included. Scales that only partially addressed religion or spirituality (e.g., with only one out of many subscales related to religion or spirituality) were excluded.

### *Screening and Selection Process*

The article search team included three members: a US faculty member in psychology originally from Taiwan and two Mainland Chinese visiting scholars in the US who had master's degrees in psychology. The articles extracted from each of the three databases (i.e., CNKI, PerioPath, PsycINFO) were reviewed independently by at least two of the three members. The reviewers screened the title and abstract of each article to determine whether the article met the inclusion/exclusion criteria. In the case of ambiguity or reviewer discrepancies, the full-texts were then obtained and analysed by the two reviewers independently to decide whether or not they fitted the inclusion/exclusion criteria. Disagreements at either the abstract or full-text screening stages were resolved via discussion among the three team members.

### *Appraisal of Quality*

Using Terwee et al.'s (2007) study as a guideline, we included the following aspects in our evaluation of each article: content validity, internal consistency, test-retest reliability, structure validity, convergent validity, and discriminant validity. For content validity, we examined whether literature reviews, previous theories, cultural factors, qualitative data, expert reviews, and pilot studies were included or conducted in the scale development process. Internal consistency was indicated by Cronbach's alpha. For test-retest reliability, we examined whether simple correlation or intraclass correlation coefficient (ICC) between two or more time points were included. To evaluate structural validity, we examined whether exploratory factor analysis (EFA) and confirmatory

factor analysis (CFA) were conducted, as well as corresponding psychometric indices, including comparative fit index (CFI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Finally, we listed whether analyses of convergent and discriminant validity or other types of validity were conducted and reported in the studies.

## Results

The initial search in the three databases resulted in 1682 potential papers, among which 1001 were from CNKI, 101 were from PerioPath, and 580 from PsycINFO. After the systematic screening process described above, we ended up including 26 articles for our review. Among these articles, nine were from CNKI, seven were from PerioPath, and ten were from PsycINFO. The complete selection process is documented in a flowchart in Figure 1.

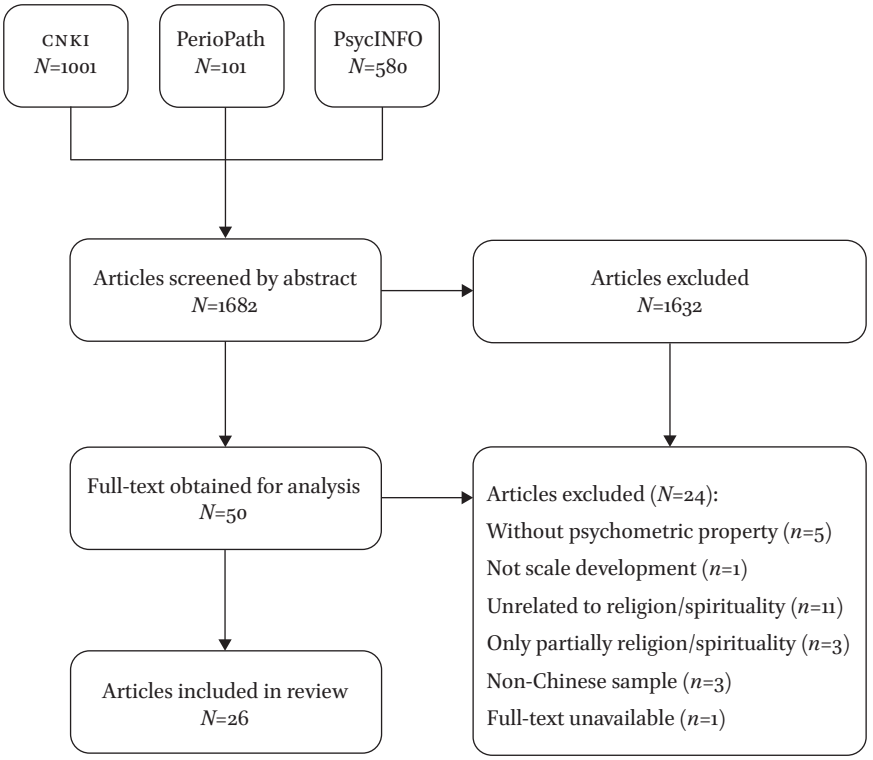



FIGURE 1 *Flowchart of article search and selection process.*

Of the 26 scales reviewed, six focused specifically on Christianity, two aimed at Buddhism, one aimed at multiple religions (i.e., Buddhism, Daoism and other indigenous Chinese religions), and the rest did not focus on any specific religion. In addition, seven studies focused exclusively on spirituality, without any specific focus on religiosity. There was some overlap among articles that examined the same scale. For example, there were five different studies involved in the translation and/or adaptation of the Duke University Religion Index (DUREL). Several of these different studies were conducted by the same group of researchers with the scale tested in different samples across China. There were also two different articles that examined the Francis Scale of Attitude toward Christianity. There was also one article that examined two scales. In short, among the 26 articles, there were only 22 unique scales examined. The descriptions of each scale, including the authors, number of items, subscales, and sample populations are listed in Table 1.

### *Assessment of Scale Development Process and Psychometric Properties*

The 26 articles were further assessed following the criteria used in Townsend-White, Pham, and Vassos's (2012) systematic review article. The assessment was first independently conducted by the second and third authors. Then the two authors checked each other's work and examined discrepancies to ensure accuracy. The assessment included two major parts—scale development process and psychometric properties. The scale development process involved the various elements to ensure content validity. The psychometric properties included indicators of reliability, factor structure, and validity. A summary of the assessment of the 26 articles is presented in Table 2.

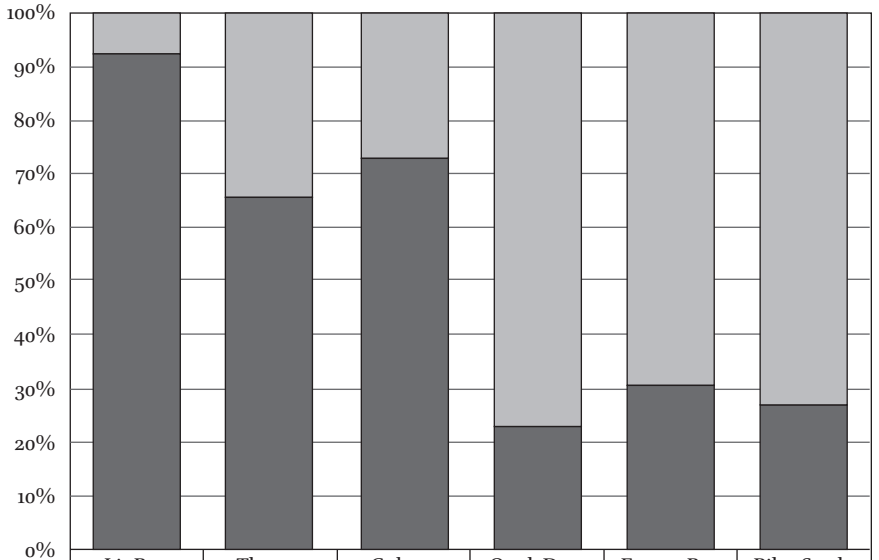
#### Content Validity

We first examined how the scale items were developed. Among the 22 unique scales included in this review, eight were original, seven were translations of English scales, five were modifications of previous scales, one was combining several previous scales, and another one did not mention the scale development process in the article. Most studies contained some level of literature review, previous theories, and cultural elements in the scale item development process. However, the level of detail provided in the descriptions varied widely. In general, the descriptions of articles published in journals from Mainland China provided less information. This is probably due to the fact that journals in Mainland China often have a rather strict word limit, which may not allow

sufficient space to provide information regarding the various elements of the item development process. Among the 26 articles, six indicated that qualitative interviews were conducted as bases for developing the scale items. Eight articles indicated that the scale items underwent a review process by experts in the topic area. Seven articles reported that pilot studies were conducted prior to the scales being tested in large samples. A summary of the percentage of articles that included each of these scale development elements is presented in Figure 2.

### Internal Consistency

With one exception (Yu, 2011), all the other 25 studies reported Cronbach's alpha as an index of internal consistency. Most scales had internal consistency of adequate magnitude (at least .60). Cronbach's alphas were provided for the total score as well as subscale scores, and for some studies, they were also provided for subsamples (e.g., different ethnic groups, different religious groups) within the study. Among the 25 studies that reported Cronbach's alpha, 11 studies (44%) had alphas that were all above .80, and 11 studies (44%) had alphas




| | Lit Rev | Theory | Culture | Qual. Data | Expert Rev | Pilot Study |
|---|---|---|---|---|---|---|
| no | 2 | 9 | 7 | 20 | 18 | 19 |
| yes | 24 | 17 | 19 | 6 | 8 | 7 |

FIGURE 2 *Summary of Scale Development Process*

TABLE 1 *Description of 22 Chinese PoR measures*

| Scale | Authors | Specific Religion[1] | Item No. | Subscale No. | Subscale Names | Spiritual only | Sample from[2] | Believer only | Sample type[3] |
|---|---|---|---|---|---|---|---|---|---|
| Age-Universal I-E Scale-12 | Yu (2011) | | 12 | 2 | Intrinsic, Extrinsic | Y | MC | | CS |
| Attribution of Religious Experience Questionnaire for College Students | Lin, Zhang et al. (2013) | | 15 | 3 | External Religious Attribution, Internal Religious Attribution, Nonreligious Attribution | | MC | | CS |
| Buddhism Relative Questionnaire | Hu et al. (2011) | B | 29 | 5 | Daily Experience, Buddhism Belief, Buddhism Copping, Buddhism Practice, Lifestyle | | MC | Y | SP |
| Buddhist Reincarnation Beliefs Scale | Hui & Coleman (2012) | B | 8 | 1 | N/A | | HK | Y | CM |
| Christian-Based Grief Recovery Scale | Pan et al. (2014) | C | 38 | 6 | Spiritual Wellbeing, Recovering Meaning & Sense of Control, Ongoing Physical & Emotional Responses, Reassuring Faith, Strongly Missing a Deceased Loved One, Life Disturbance | | TW | Y | CS, SP |

| Scale | Authors | Specific Religion[1] | Item No. | Subscale No. | Subscale Names | Spiritual only | Sample from[2] | Believer only | Sample type[3] |
|---|---|---|---|---|---|---|---|---|---|
| Daily Spiritual Experience Scale | Ng et al. (2009) | | 16 | 1 | N/A | Y | HK | | SP |
| Demands of Spiritual Caring Education Scale | Yang & Huang (2006) | | 24 | 3 | Spirituality Improving Activity, Spiritual Caring Related Knowledge, Application of Spiritual Caring | Y | TW | | SP |
| Duke University Religion Index | Chen et al. (2014); Liu & Koenig (2013); Wang et al. (2013); Wang, Rong et al. (2014); Wang, Wang et al. (2014) | | 5 | 3 | Organizational Religious Activity, Non-Organizational Religious Activity, Intrinsic Religiosity | | MC, HK | Y & N | CM, CS, SP |
| Faith Maturity Scale | Hui et al. (2011) | C | 12 | 2 | Vertical, Horizontal | | MC, HK | Y | CM |
| Francis Scale of Attitude toward Christianity | Francis et al. (2002); Tiliopoulos et al. (2013) | C | 24 | 1 | N/A | | HK, MC | Y & N | CM, SP |
| Hoge Intrinsic Religiosity Scale | Liu & Koenig (2013) | | 10 | 2 | Extrinsic Religiosity, Intrinsic Religiosity | | MC | | SP |

TABLE 1 *Description of 22 Chinese PoR measures* (cont.)

| Scale | Authors | Specific Religion[1] | Item No. | Subscale No. | Subscale Names | Spiritual only | Sample from[2] | Believer only | Sample type[3] |
|---|---|---|---|---|---|---|---|---|---|
| Hood's Mysticism Scale | Chen et al. (2012) | C | 32 | 8 | Timelessness–Spacelessness,<br>Ego Loss,<br>Ineffability,<br>Inner Subjectivity,<br>Unity,<br>Positive Affect,<br>Sacredness,<br>Noetic Quality | | MC, HK | | CM |
| Native Afterlife Beliefs Scale for Undergraduates | Tsai & Ou (2007) | | 70 | 12 | Level of Belief,<br>Deciding Factors<br>After Life Conditions | Y | TW | | CS |
| Personal Spiritual Needs Scale | Hu & Wong (2013) | | 15 | 3 | Constructive,<br>Affective,<br>Divine,<br>Meditative | Y | TW | | SP |
| Questionnaire on Triggers of Religious Experiences in College Students | Song et al. (2015) | | N/A | 5 | Stressful Circumstances,<br>Personal Ritual,<br>Aesthetic & Creativity,<br>Group Ritual,<br>Music | | MC | | CS |

| Scale | Authors | Specific Religion[1] | Item No. | Subscale No. | Subscale Names | Spiritual only | Sample from[2] | Believer only | Sample type[3] |
|---|---|---|---|---|---|---|---|---|---|
| Religious Behavior Instrument | Yu & Huang (2009) | C | 8 | 3 | Individual Praying,<br>Church Attendance,<br>Religious Activity of Family Members | | TW | Y | SP |
| Religious Behavior Questionnaire | Lin, Song et al. (2013) | | 29 | 6 | Learning Knowledge,<br>Praying,<br>Worshiping,<br>Geomantic Omen,<br>Fortune Telling,<br>Religious Taboo | | MC | | CS |
| Religious Belief Scale | Su & Liu (2011) | B, T, O | 16 | 4 | Qi,<br>Soul,<br>Karma,<br>Divine Intervention | | TW | | CM |
| Scale of Spiritual Health of Senior High School Teachers | Chang & Chen (2008) | | 33 | 5 | Overcoming Obstacles,<br>Connection with Others,<br>Meaning in Life,<br>Religious Hope,<br>In Awe of Nature | Y | TW | | SP |

TABLE 1 *Description of 22 Chinese PoR measures* (cont.)

| Scale | Authors | Specific Religion[1] | Item No. | Subscale No. | Subscale Names | Spiritual only | Sample from[2] | Believer only | Sample type[3] |
|---|---|---|---|---|---|---|---|---|---|
| Spiritual Care Attitude Scale | Chiang et al. (2014) | | 15 | 3 | Spiritual Growth, Core Value, Spiritual Care | Y | TW | | SP |
| Spiritual Health Scale Short Form | Hsiao et al. (2013) | | 24 | 5 | Connection to Others, Meaning Derived from Living, Transcendence, Religious Attachment, Self-Understanding | Y | TW | | SP |
| Spiritual Questionnaire for College Students | Fu & Ding (2014) | | 21 | 5 | Love & Compassion, Daily Delights, Meaning in Life, Transcendental Experience, Social Belongingness | Y | MC | | CS |

*Note*: [1]C = Christianity, B = Buddhism, T = Taoism, O = Others. [2]MC = Mainland China, HK = Hong Kong, TW = Taiwan. [3]CS = College Students, CM = Community Sample, SP = Special Sample.

TABLE 2 *Evaluation of 26 articles on Chinese PoR measures*

| Authors | Scale Development & Content Validity | | | | | | | Reliability | | Structural Validity | | | | | | Construct Validity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Item Dev[1] | Lit Rev | Theory | Culture | Qual. Data[2] | Expert Review | Pilot Study | Alpha[3] | Test-retest[4] | EFA | CFA | CFI[5] | RMSEA[6] | SRMR[7] | Con | Dis | Other |
| Chang & Chen (2008) | C | Y | Y | Y | | Y | Y | H | | Y | | | | | | | Y |
| Chen et al. (2012) | T | Y | Y | Y | | | | H,M | | | Y | M | H | H | | | Y |
| Chen et al. (2014) | T | Y | | | | | Y | H,M | H* | Y | Y | H | H | | | | |
| Chiang et al. (2014) | O | Y | Y | | I | Y | | H | | Y | Y | H | M | | Y | | Y |
| Francis et al. (2002) | T | Y | Y | Y | | | | H | | Y | | | | | Y | | Y |
| Fu & Ding (2014) | O | Y | Y | Y | I | Y | | H,M,L | | Y | Y | M | H | | | | Y |
| Hsiao et al. (2013) | M | Y | Y | | I | | | H | | | Y | H,M | M | H | Y | Y | Y |
| Hu & Wong (2013) | O | Y | Y | | | | | H | | Y | Y | M | M | H | Y | Y | Y |
| Hu et al. (2011) | M | | | Y | | Y | | H,M | H | Y | | | | | Y | | |
| Hui & Coleman (2012) | O | Y | Y | Y | | Y | | H | | Y | | | | | Y | | |
| Hui et al. (2011) | T | Y | Y | Y | | | | H,M | | | Y | M | M | | Y | | Y |
| Lin, Song et al. (2013) | O | Y | | Y | | | | H,M | | | Y | M | M | | | | Y |
| Lin, Zhang et al. (2013) | O | Y | | Y | | | | H,M | | | Y | L | M | | | | Y |
| Liu & Koenig (2013) | M | Y | Y | Y | | | | H,M | | Y | | | | | Y | | |

TABLE 2 *Evaluation of 26 articles on Chinese PoR measures* (cont.)

| Authors | Scale Development & Content Validity | | | | | | | Reliability | | Structural Validity | | | | | Construct Validity | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Item Dev[1] | Lit Rev | Theory | Culture | Qual. Data[2] | Expert Review | Pilot Study | Alpha[3] | Test-retest[4] | EFA | CFA | CFI[5] | RMSEA[6] | SRMR[7] | Con | Dis | Other |
| Ng et al. (2009) | T | Y | | Y | | | Y | H | | Y | | | | | Y | Y | |
| Pan et al. (2014) | O | Y | Y | Y | I | Y | Y | H,M | | Y | | | | | Y | | |
| Song et al. (2015) | O | Y | | Y | I | | | H | | Y | Y | | M | | | | Y |
| Su & Liu (2011) | O | Y | Y | Y | | | | H,M | | Y | | | | | | | Y |
| Tiliopoulos et al. (2013) | T | Y | Y | | | | | H | | Y | | | | | Y | | |
| Tsai & Ou (2007) | O | Y | Y | Y | I | | Y | H,M | H,M,L | Y | | | | | | Y | Y |
| Wang et al. (2013) | T | | | | | Y | Y | H,M | H,M,L* | | Y | | | | | | |
| Wang, Rong et al. (2014) | T | Y | Y | Y | | | Y | H,M,L | H,M* | Y | | | | | | | Y |
| Wang, Wang et al. (2014) | T | Y | | Y | | | | H,M,L | H,M* | | Y | | | | | | Y |
| Yang & Huang (2006) | M | Y | Y | | | Y | | H | | Y | | | | | | | Y |
| Yu (2011) | T | Y | | Y | | | | | L | | Y | H | M | | | | Y |
| Yu & Huang (2009) | M | Y | Y | Y | | | | H | | Y | | | | | | | Y |

*Note*: * Presented as ICC. [1]T = Translated, C = Combined, M = Modified, O = Original. [2]I = Interview. [3]L < .60 < M < .80 < H. [4]L < .70 < M < .80 < H for correlation; L < .30 < M < .60 < H for ICC. [5]L < .90 < M < .95 < H. [6]H < .05 < M < .10 < L. [7]H < .08 < L.

that were at least .60. For the three studies with Cronbach's alphas that ranged from low to high, they were mostly due to the studies examining alphas for all subscales across various subsamples.

### Test-retest Reliability

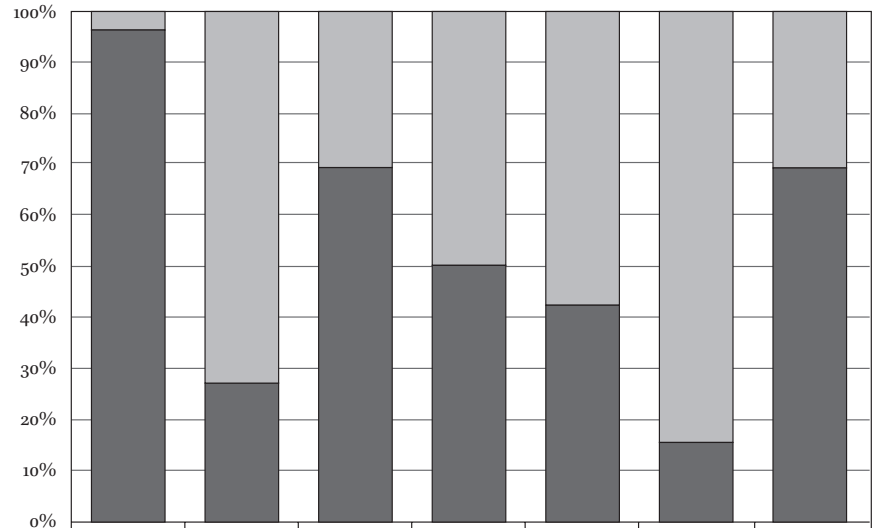
Seven studies conducted retests with a complete or a partial sample, three of which were conducted two weeks after the initial test, two were conducted one week later, one was conducted four weeks later, and another one was conducted an average of 2.5 days later. Four studies reported ICC as an index for test-retest reliability and the remaining studies reported Pearson's correlation. Among the four studies that reported ICCs, one (25%) had an ICC value above .60, and three (75%) had mixed results across subscales. Among the three studies that reported correlations, one (33%) was above .80, one (33%) was below .70, and one (33%) had mixed results across subscales.

### Factor Structure

All the studies reviewed contained some assessment of the factor structure. Thirteen studies (50%) only conducted EFA, eight studies (31%) only conducted CFA, and five studies (19%) conducted both. The scales reviewed all showed acceptable levels of CFA indices. Among the studies that conducted both EFA and CFA, most used a different sample for the CFA to cross-validate the EFA factor structure. However, some studies conducted the EFA and CFA using the same sample, or did not specify whether the samples were the same or not.

### Construct Validity

The validity of scales was examined through various methods across the studies. Some studies explicitly examined convergent and discriminant validities, whereas most others reported mean comparisons on the scale/subscale scores across groups. Eleven studies (42%) examined convergent validity by demonstrating that the scale scores correlated with related constructs in the expected direction. Four studies (15%) examined discriminant validity. One study only examined discriminant validity. Sixteen studies (62%) compared mean scores of the given scale across different groups (e.g., demographic variables, single item questions). One study examined incremental validity. Two studies (8%) reported neither correlation nor mean score group comparison. A summary of the percentage of articles that examined each of these psychometric properties indices is presented in Figure 3.




| | Cronbach Alpha | Test-Retest | EFA | CFA | Convergent Validity | Discriminant Validity | Other Validity |
|---|---|---|---|---|---|---|---|
| no | 1 | 19 | 8 | 13 | 15 | 22 | 8 |
| yes | 25 | 7 | 18 | 13 | 11 | 4 | 18 |

FIGURE 3 *Summary of Psychometric Properties reported*

## Discussion

In this paper, we reviewed scale development and psychometric evaluation studies in the field of psychology of religion in Chinese cultures. What we found was both promising and concerning. First of all, though the studies and scales were relatively small in number, there was a rapid increase during recent years. All the studies we reviewed were published after 2001, and 20 out of 26 studies were published after 2010. Although not included in the current review, we identified 40 master's theses and doctoral dissertations on this topic when we searched through the literature. As a result, we believe the field is still in its infancy and bears great potential. We expect to see more studies published in this area within the next decade. In our review, we also found that nearly forty percent of the scales were originally developed by Chinese scholars and that this was a growing trend. As suggested by Yang (1999), indigenous scales may better capture the unique characteristics of local culture. Although the majority of the scales were still either a direct translation or a modification of some Western scale, it is encouraging to see the trend of

indigenous studies and we encourage Chinese psychologists to continue this trend in future studies.

The scales we reviewed covered a wide range of topics, including general measures of religiosity (Wang, Sun, Rong, Zhang, & Wang 2013; Wang, Wang, Rong, Sun, & Zhang, 2014; Chen, Wang, Michael, Sun, & Cheng, 2014; Liu & Koenig, 2013; Yu, 2011; Hu, Huang, Huang, Zeng, Mei, Li, & Huang, 2011), spirituality (Ng, Fong, Tsui, Au-Yeung, & Law, 2009; Hu & Wong, 2013; Fu & Ding, 2014), religious experience (Lin, Zhang et al., 2013; Song et al., 2015), religious belief (Tsai & Ou, 2007; Su & Liu, 2011; Hui & Coleman, 2012), religious behaviour (Yu & Huang, 2009; Lin, Song, & Zhang, 2013), spiritual health (Pan, Deng, Tsai, Chen, & Yuan, 2014; Yang & Huang, 2006; Chang & Chen, 2008 Hsiao, Chiang, Lee, & Han, 2013), faith maturity (Hui, Wai Ng, Ying Mok, Ying Lau, & Cheung, 2011), attitude toward Christianity (Francis, Lewis, & Ng, 2002; Tiliopoulos, Francis, & Jiang, 2013) and mysticism (Chen, Zhang, Hood, & Watson, 2012). There were some broad topics missing in the Chinese literature. Several related measures on topics such as religious/spiritual development and religious struggle and strain were developed in Western studies of psychology of religion (Hill & Edwards, 2013), which could be considered for future Chinese studies.

While most studies did not aim at a specific religion, two measures were only applicable to Christianity and another two only applicable to Buddhism. Though general measures could be useful, measures designed for a specific religion may better describe the specific beliefs and practices of a given religion. Due to the diversity of religions held by the Chinese compared to the US (Du, 2010), future studies should focus on other Chinese religions as well, such as Islam, Daoism, Tibetan Buddhism, and folk beliefs.

Furthermore, of all the 26 studies only seven studies used college student samples while the remaining studies examined their measures in community samples or specific groups (e.g., nurses, church attenders). We consider this a good sign, as college students are often considered to be WEIRDer (Western, Educated, Industrialized, Rich, Democratic; Henrich, Heine & Norenzayan, 2010) than the general population and may likely lead to biased results. Seven studies developed their scales among religious believers, which is especially important for religious scales. We encourage researchers to use diverse groups of religious believers, especially community religious believers as well as exemplary figures of varying religious traditions.

Through our examination of the scale development process, we found that although some studies provided detailed descriptions of the development process, others were brief and vague in this regard, especially for articles published in Mainland China. Some studies only provided a few statements to describe

the generation and screening of items. In addition, the descriptions of expert review and interview processes were often lacking in detail, without a clear link to their contributions. Some studies did not provide a thorough literature review nor a strong theoretical framework. In Hill and Edwards' (2013) words, "good theory and good measurement go hand in hand" (p. 53). We believe a good scale must have a solid theoretical background and urge researchers to focus more on theories before rushing into scale development.

As far as psychometric properties of the articles reviewed are concerned, a majority of the studies utilized EFA to select items. However, only half of the studies incorporated CFA to cross-validate the factor structure. In general, most of the published scales reported adequate psychometric statistics though some did not report sufficient information. The scales all showed acceptable internal consistencies. In addition, the scales had good test-retest reliability and structure validities, but only for those that reported these indices. However, half of the studies did not test the scale against any external criterion, which made their construct validity vulnerable. It is also worth noting that in this review we excluded studies that reported internal consistency only. Such studies were uncommon but were found in our search. As a result, we suggest that researchers should not only examine the scale's factor structure but also examine its relationship with other measures whenever possible. Suggestions on best practices for finalizing scale items and examining psychometric properties can also be found in the last section of this paper.

### *Best Practice for Scale Development*

In this section, we will describe scale development steps to offer recommendations on best practices for developing culturally sensitive measures. We will provide examples from the studies that we reviewed to illustrate best practices. We also provided a checklist of several key points of scale development in Table 3.

#### Immersive Understanding of Religious Tradition

Due to the wide diversity of religious traditions and the overall low percentage of religious believers, it is not uncommon for PR researchers in China to have little first-hand knowledge of the religious tradition that they are studying. Therefore, for those that are less familiar with the religious tradition of the groups they are researching, a first task is to immerse oneself in the religious culture like an ethnographer. This would mean attending religious services, obtaining an informant, reading the religious literature, holding focus groups, and interviewing persons in the tradition on the construct of interest.

TABLE 3 *Checklist of key points in scale development*

| | |
|---|---|
| 1. | Obtain firsthand knowledge of the religious tradition of interest. |
| 2. | Conceptualize the construct of interest with an operational definition and a clear scope. |
| 3. | Conduct culturally relevant literature review utilizing both classic and contemporary theoretical and empirical literature. |
| 4. | Use qualitative methods, such as in-depth interviews, focus groups, and open question surveys, when needed, to inform item development. |
| 5. | Use an iterative process including literature review, qualitative data collection, expert review, and pilot testing to finalize the scale item pool. |
| 6. | Seek input from experts of various sources to generate and revise the scale items and rating format. |
| 7. | Collect two datasets from target population, one for exploratory factor analysis and the other for confirmatory factor analysis. |
| 8. | Conduct exploratory factor analysis (EFA) to finalize the scale items and optimize the length of the scale in a first sample. |
| 9. | Conduct confirmatory factor analysis (CFA) to cross-validate the factor structure yielded through the EFA in a different sample. |
| 10. | Examine and report psychometric properties, including indices for reliability, construct validity, and incremental validity. |

### Conceptualize Construct of Interest

An initial step of a scale development project is to conceptualize and operationalize the construct of interest. This step involves clearly defining the construct which the scale will be measuring. In addition to a conceptual definition, it is also important to translate the construct into an operational definition (Heppner et al., 2015). This process also includes defining the scope of the scale, such as what will and will not be measured by the scale as well as the type of setting and population in which this particular scale will be used. As part of the conceptualization and operationalization process, researchers should also determine the content domains that fall under the construct. Conceptualizing constructs is often an iterative process, conducted with reviewing the literature and/or collecting qualitative data. For example, in Tsai and Ou's (2007) process of developing a Chinese indigenous Afterlife Beliefs Scale, they paid careful attention to defining the construct. They started by utilizing both Chinese and English dictionary definitions to first come up with

a definition for the term 'afterlife beliefs'. Afterwards, they followed with an operational definition that included twelve dimensions.

### Conduct Culturally Relevant Literature Review

A critical issue in developing culturally indigenous scales pertains to the literature review that informs the item development process. It would be important that the theoretical and empirical research literature reviewed is culturally relevant. If available, researchers should review a broad range of sources that includes classical Chinese literature, contemporary Chinese literature, as well as Chinese research literature. In terms of classical Chinese literature, researchers could consult classical writings, autobiographies, and religious texts to gather information on traditional Chinese values and concepts. Due to industrialization and modernization, Chinese society has gone through rapid cultural changes during the past several decades (e.g., New Cultural Movement and Cultural Revolution). Therefore, it is also important to utilize contemporary Chinese literature (e.g., contemporary popular books, religious materials, and online resources) to gather information of thoughts and concepts more closely aligned to the modern-day people. Finally, it is important to utilize both theoretical and empirical research literature that is applicable to the Chinese population. We suggest that it is most helpful to consult theories developed with and for the Chinese population. However, the current research literature is predominantly based on Western theories and findings with Western samples. Although Western literature as well as cross-cultural studies can be helpful, caution should be taken when using these resources (Yang, 1999). To illustrate the literature review process, Tsai and Ou (2007) first reviewed the Chinese classical literature for a traditional cultural perspective. They reviewed Confucius and Daoist writings in order to gain an understanding of the nature of afterlife beliefs prior to Buddhism's influence on China. They also analysed Buddhist writings after Buddhism was introduced into China during the East Han dynasty. In addition to the Chinese literature, Tsai and Ou consulted Western scales that measured afterlife beliefs, and incorporated some of the concepts to construct their own model of the afterlife beliefs.

### Qualitative Interviews and Open Questions

As mentioned above, Chinese theories and research on psychology of religion are very limited. Therefore, collecting qualitative data to inform scale development would be a useful method to compensate for the lack of developed theories (Yang, 1999). There are different ways of collecting qualitative data, including in-depth interviews, focus groups, and open question surveys. There are advantages for each of these three different methods. Interviews can pro-

vide more in-depth information based on the person's individual experiences and thoughts. Data collected from this method is often richer in depth of content. Using a focus group has the advantage of stimulating ideas via brainstorming and group discussion. The advantage of using open question surveys is the ability to reach a broader sample of individuals in a more efficient way. Therefore, this method provides data with more breadth. Researchers can use whichever method that aligns with their goals, or a combination of different methods to gather a comprehensive set of data. Conducting interviews was method that Tsai and Ou (2007) incorporated in conjunction with Chinese literature and Western studies to develop the items of their Afterlife Beliefs Scale. Through these multiple sources, the researchers came up with three higher order domains for afterlife beliefs—Level of Belief, Deciding Factors, and After Life Conditions. They argued for the importance of learning about afterlife through interviews from the Chinese population in which the scale would be developed to be used with.

### Generate Items, Indicators, and Response Formats

To generate scale items, researchers should start with developing a pool of items which is often 3 to 4 times the target number of items for the final scale (DeVellis, 2012). Items should be developed based on literature review as well as data collected from qualitative methods. In addition to items generated by the research team, we suggest also seeking input from others that have experiences or expertise related to the construct of the scale. It would be most helpful to seek input from a variety of resources, including religious leaders, academic scholars, and practitioners. For example, in developing a Buddhism scale, it would be helpful to consult with Zen masters, Buddhist scholars, Buddhist monks, and common Buddhist believers. Different people may provide different perspectives on the construct of interest.

After gathering information and ideas for the construct, the concepts need to be translated into scale items. DeVellis (2012) provided some concrete guidelines around 'do's and don'ts' for writing scale items. He suggested that item statements should be brief, precise, easy to read, have only one central thought, phrased in a positive language, and all worded in the same direction. On the other hand, DeVellis urged to avoid awkward wording, ambiguous pronouns, ambiguity in meaning, double negatives, double-barrelled items (i.e., items that ask multiple question), lengthy items, irrelevant information, extreme terms (e.g., all or nothing), indeterminate terms (e.g., sometime, frequently), and reversed items. It is also important to decide on using the response format (e.g., Likert scale, yes-no checkbox, semantic differential) and indicators (e.g., degree of agreement, frequency, likelihood) that are most appropriate for the aims of the scale.

### Content Analysis and Pilot Testing

After items are developed and polished, it is important to obtain input from experts in the field. For example, it would be useful to get expert feedback on whether the items are appropriate in measuring the construct of interest and whether certain aspects of the construct are not represented by the items. More specifically, the reviewer can be asked to indicate the level of item-construct match, sort items with construct domains, revise and edit item statements, and suggest additional items (Worthington & Whittaker, 2006). It is also important to go through the process of pilot testing and revising to refine the scale items before administering the item pool to a large sample of participants (Heppner et al., 2015). During this process the goal is to obtain feedback regarding the clarity of items. In addition, the instructions and response format should also be adjusted to most accurately fit the item statements. It would also be important to ensure that items are written at the appropriate reading level of the target population (DeVellis, 2012). Among the studies that were reviewed, Chiang and colleagues' (2014) Spiritual Care Attitude Scale was a study that went through the expert review process. The authors sought feedback from seven scholars and clinical experts with different expertise (e.g., qualitative methods, oncology nursing, and hospice care) to establish content validity. The items and factors were reviewed and revised accordingly. Two items were removed; one due to redundancy, and the other due to being clinically irrelevant. Content validity was ensured prior to collecting large scale data.

### Sample and Collect Data

After finalizing the item pool for the scale, the next step would be to collect data from a large sample of participants. The target sample of participants should align with the population that the scale is intended to be used with. In other words, the sampling method will impact the generalizability of the scale. Ideally, two samples should be collected, in which one would be used for exploratory factor analysis and the other used for confirmatory factor analysis (Cabrera-Nguyen, 2010). For each sample, the sample size should be at least five (Gorsuch, 1983) to ten (Nunnally, 1978) times the size of the number of items, with at least 100 participants (Gorsuch, 1983). For example, if there were 40 items in the item pool, a sample size of at least 200 participants, and ideally 400 participants, would be appropriate.

### Finalize Items and Optimize Scale Length

The first sample should be used to conduct exploratory factor analysis (EFA). EFA will be used to determine the number of factors/domains of the scale and to select the most appropriate items from the item pool. Based on the item

descriptions under each factor, researchers will label factors accordingly. For more details on EFA, we encourage reading Cabrera-Nguyen (2010) and Worthington and Whittaker (2006).

A second sample should be used to conduct confirmatory factor analysis (CFA). It is important to use different datasets when conducting EFA and CFA (Hurley et al., 1997). The goal of CFA is to cross-examine the factor structure yielded through EFA with an independent dataset. Fit indices are used to determine whether the data fits the proposed factor structure. In addition, multi-group confirmatory factor analysis can be used to examine whether the factor structure is invariant across different groups (e.g., gender, religious orientation). For more details on CFA, we encourage reading Brown (2015), Kline (2015) and Worthington and Whittaker (2006).

Optimizing the scale length can be done by removing redundant items (DeVellis, 2012). The shorter the scale, the easier it is to administer. Redundant items can be identified through the modification indices that suggest overlapping error variances (Wang, Wei, Zhao, Chuang, & Li, 2015). High internal consistency reliability coefficients (i.e., alphas over .90) are also indicators of item redundancy (DeVellis, 2012). When determining which items to be removed, it would be helpful to utilize results from item analysis. Item-scale correlation (higher), item mean (closer to centre), and item variance (higher) can be used to determine better performing items (DeVellis, 2012). Chiang et al.'s (2014) study was a good example of conducting both EFA and CFA to examine and validate the factor structure in two different samples. Through the EFA process they determined the factors of their Spiritual Care Attitude Scale—Spiritual Growth, Core Value, and Spiritual Care. They eliminated items through both the EFA and CFA processes, and ended up with a 15-item scale that started from a 25-item pool.

### Examine Psychometric Properties

After the factor structure and final scale items are determined, the psychometric properties of the scale should be examined and reported. First, Cronbach's alpha, which measures internal consistency reliability, should be reported. Test-retest reliability, which represents stability over time, is another often used reliability indicator. Second, construct validity of the scale should be examined and reported. The construct of interest is first hypothesized to correlate positively or negatively with other constructs based on theory, and then examined with data. One form of construct validity is convergent validity, which suggests that the scale scores should correlate positively with measures of similar constructs. Another form of validity is discriminant validity, which suggests that the scale scores should not be significantly correlated with measures of unrelated constructs. Social desirability is a commonly used variable

to demonstrate discriminant validity. Moreover, examining incremental validity would also be helpful, especially for cultural applicability. Incremental validity can support the newly developed scale being an improvement over existing scales by predicting variances above and beyond the existing scales. In terms of examining psychometric properties, Chiang et al.'s (2014) study of the Spiritual Care Attitude Scale was a good example. They reported indicators of convergent validity as well as concurrent validity. They demonstrated that the Chinese indigenous Spiritual Care Attitude Scale was moderately correlated with a Spiritual Health Scale-Short Form.

## Conclusion

Psychology of religion research in China is still in its infancy, so the scale development in this field is lagging behind. However, scales are important tools for scientific research. Through this literature review, we hoped not only to provide Chinese PR researchers with an overview of the measurement landscape, but also to facilitate the process of identifying appropriate scales efficiently. We hope to see more Chinese indigenous PR scales developed with rigor and quality to stimulate the future growth of this field.

## References

Brown, T. A. (2015). Confirmatory factor analysis for applied research. New York: Guilford Press.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research, 1*, 99–103. doi:10.5243/jsswr.2010.8.

Chang, S. M., & Chen, H. T. (2008). A study on spiritual health and its related factors of the senior high school teachers in Kaohsiung area [高雄地區高中教師靈性健康及其相關因素之研究]. *Journal of Life-and-Death Studies*, 7, 89–138.

Chen, H. H., Wang Z. Z., Michael R. P., Sun Y. L., & Cheng, H. G. (2014). Internal consistency and test-retest reliability of the Chinese version of the 5-item Duke University Religion Index [中文版五条目杜克大学宗教指数量表的内部一致性和重测信度]. *Shanghai Archives of Psychiatry, 26*, 300–309.

Chen, Z., Zhang, Y., Hood, R. W., & Watson, P. J. (2012). Mysticism in Chinese Christians and non-Christians: Measurement invariance of the Mysticism Scale and implications for the mean differences. *International Journal for the Psychology of Religion, 22*, 155–168. doi:10.1080/10508619.2011.638586.

Chiang, Y. C., Chiang, H. Y., Lee, H. C., Han, C. Y., & Hsiao, Y. C. (2014). The construction and evaluation of a Spiritual Care Attitude Scale [靈性照護態度量表建構與信效度檢定]. *Journal of Nursing and Healthcare Research, 10*, 102–112. doi:10.6225/JNHR.10.2.102.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). London: Sage publications.

Du, Y. F. (2010). Current situation and development trend of religions in contemporary China [当代中国宗教发展的现状及趋势]. *Journal of the Central Institute of Socialism, 5*, 77–83.

Dueck, A., & Han, B. (2012). Psychology of religion in China. *Pastoral Psychology 61*, 605–622. doi:10.1007/s11089-012-0488-2.

Francis, L. J., Lewis, C. A., & Ng, P. (2002). Assessing attitude toward Christianity among Chinese speaking adolescents in Hong Kong: The Francis scale. *North American Journal of Psychology, 4*, 431–439.

Fu, L. H., & Ding, J. L. (2014). An investigation on the state of college students spirituality [大学生灵性现状调查研究]. *Journal of Hechai University, 34*, 123–128.

Gallup Inc. (2015). "*Religion*". Retrieved from http://www.gallup.com/poll/1690/religion.aspx.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83. doi:10.1017/s0140525x0999152x.

Heppner, P. P., Wampold, B. E., Owen, J. J., Thompson, M. N., & Wang, K. T. (2015). *Research Design in Counseling* (4th ed.). Belmont, CA: Thomas Brooks/Cole.

Hill, P. C., & Edwards, E. (2013). Measurement in the psychology of religiousness and spirituality: Existing measures and new frontiers. In K. I. Pargament, J. J. Exline, & J. W. Jones (Eds.), *APA handbook of psychology, religion, and spirituality* (*Vol 1*): *Context, theory, and research. APA handbooks in psychology* (pp. 51–77). Washington, DC: APA.

Hill, P. C., & Hood, R. W. (Eds.). (1999). *Measures of religiosity*. Religious Education Press.

Hsiao, Y. C., Chiang, Y. C., Lee, H. C., & Han, C. Y. (2013). Psychometric testing of the properties of the Spiritual Health Scale Short Form. *Journal of Clinical Nursing, 22*, 2981–2990. doi: 10.1111/jocn.12410.

Hu, J. S., & Wong, H. R. (2013). The effect of personal spiritual needs on organizational spirituality adoption [個人靈性需求對組織靈性接納之影響]. *Fu Jen Management Review, 20*, 127–148.

Hu. H. Y., Huang X. M., Huang P., Zeng Q. M., Mei F., Li J., & Huang Y. Q. (2011). Pilot evaluation of reliability and validity of the Buddhism Relative Questionnaire [佛教相关问卷的信度和效度试测]. *Journal of International Psychiatry, 38*, 213–216. doi: 10.13479/j.cnki.jip.2011.04.001.

Hui, C. H., Wai Ng, E. C., Ying Mok, D. S., Ying Lau, E. Y., & Cheung, S. F. (2011). "Faith Maturity Scale" for Chinese: A revision and construct validation. *International Journal for the Psychology of Religion, 21*, 308–322. doi:10.1080/10508619.2011.607417.

Hui, V. K. Y., & Coleman, P. G. (2012). Do reincarnation beliefs protect older adult Chinese Buddhists against personal death anxiety? *Death Studies, 36*, 949–958. doi:10.1080/07481187.2011.617490.

Hurley, A. E., Scandura, T. A., Schriesheim, C. A., Brannick, M. T., Seers, A., Vandenberg, R. J., & Williams, L. J. (1997). Exploratory and confirmatory factor analysis: Guidelines, issues, and alternatives. *Journal of Organizational Behavior, 18*, 667–683. doi:10.1002/(SICI)1099-1379(199711)18:6<667::AID-JOB874>3.0.CO;2-T.

Kline, R. B. (2015). Principles and practice of structural equation modeling. New York: Guilford Press.

Lin Y. L., Song, X. C., & Zhang, Y. (2013). Current situation of religious behavior among undergraduate student [大学生宗教行为现状研究-以泉州大学生为例]. *Journal of Longyan University, 31*, 62–66.

Lin, Y. L., Zhang, Y., & Song, X. C. (2013). Research on current situation of the attribution of religious experience in undergraduate [大学生宗教经验归因的现状研究-以泉州大学生为例]. *Journal of Qinghai Normal University* (*Philosophy and Social Sciences*)*, 35*, 125–128. doi:10.16229/j.cnki.issn1000-5102.2013.05.040.

Liu, E. Y., & Koenig, H. G. (2013). Measuring intrinsic religiosity: scales for use in mental health studies in China—a research report. *Mental Health, Religion & Culture, 16*, 215–224. doi:10.1080/13674676.2012.672404.

Ng, S. M., Fong, T. C., Tsui, E. Y., Au-Yeung, F. S., & Law, S. K. (2009). Validation of the Chinese version of Underwood's Daily Spiritual Experience Scale—transcending cultural boundaries? *International Journal of Behavioral Medicine, 16*, 91–97. doi:10.1007/s12529-009-9045-5.

Nunnally, J. C. (1978). *Psychometric theory* (2nd Ed.). New York: McGraw-Hill.

Pan, P. J. D., Deng, L. Y. F., Tsai, S. L., Chen, H. Y. J., & Yuan, S. S. J. (2014). Development and validation of a Christian-based Grief Recovery Scale. *British Journal of Guidance & Counselling, 42*, 99–114. doi:10.1080/03069885.2013.852158.

Singelis, T. M. (1994). The measurement of independent and interdependent self-construals. *Personality and Social Psychology Bulletin, 20*, 580–591. doi:10.1177/0146167294205014.

Song, X. C., Liu, H. Y., & Sun, J. J. (2015). Comparison of triggers of religious experiences among college students of different grades and majors [不同年级、专业大学生宗教经验触发因素的比较]. *Journal of Lishui University, 37*, 117–123. doi:10.3969/j.issn.2095-3801.2015.04.019.

Su, B. G., & Liu, Y. J. (2011). The Development of Religious Belief Scale [台灣宗教信念量表之發展]. *Taiwan Journal of Hospice Palliative Care, 16*, 168–190. doi:10.6537/TJHPC.2011.16(2).3.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., … & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*, 34–42. doi:10.1016/j.jclinepi.2006.03.012.

Tiliopoulos, N., Francis, L. J., & Jiang, Y. (2013). The Chinese translation of the Francis Scale of Attitude toward Christianity: Factor structure, internal consistency reliability, and construct validity among protestant Christians in Shanghai. *Pastoral Psychology, 62*, 75–79. doi: 10.1007/s11089-012-0457-9.

Tocqueville, A. (2003). *Democracy in America and two essays on America* (G. Bevan, Trans.). New York, NY: Penguin Group. (Original work published in 1835).

Townsend-White, C., Pham, A. N. T., & Vassos, M. V. (2012). Review: A systematic review of quality of life measures for people with intellectual disabilities and challenging behaviours. *Journal of Intellectual Disability Research, 56*(3), 270–284. doi:10.1111/j.1365-2788.2011.01427.x.

Tsai, M. C., & Ou, H. M. (2007). The construction and development of Native Afterlife Beliefs Scale for undergraduate [本土化大學生來生信念量表的建構與發展]. *Journal of Life-and-Death Studies 7*, 7–88.

van de Vijver, F. J. R., & Leung, K. (1997). Methods and data analysis for cross-cultural research. Thousand Oaks, CA: Sage.

Wang, J., Sun, Y. L., Rong, Y., Zhang, Y. H., & Wang, Z. Z. (2013). The reliability of Chinese version DUREL Scale [DUREL 量表的中文翻译及其信度评价]. *Journal of Ningxia Medical University, 35*, 550–552.

Wang, K. T., Wei, M., Zhao, R., Chuang, C. C., & Li, F. (2015). The Cross-Cultural Loss Scale: Development and psychometric evaluation. *Psychological Assessment, 27*, 42–53. doi:10.1037/pas0000027.

Wang, Z. Z., Wang, J., Rong, Y., Sun, Y. L., & Zhang, Y. H. (2014). The reliability of Chinese version DUREL Scale in rural residents [中文版杜克信仰量表在回汉族农村居民中的应用]. *Journal of Ningxia Medical University, 36*, 990–992.

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*, 427–443. doi:10.2466/08.17.PR0.115c19z8.

WIN-Gallup International (2012). "*Global index of religiosity and atheism*". Retrieved from http://www.wingia.com/web/files/news/14/file/14.pdf.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research a content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806–838. doi:10.1177/0011000006288127.

Wu, J. (2007). *Religious believers thrice the estimate*. China Daily, 2010-11-19.

Yang, K. P., & Huang, C. K. (2006). The Demands of spiritual care education among faculties and students in a nursing college [護理專科學校師生靈性關懷教育需求之探討]. *Journal of Cardinal Tien College of Nursing, 4*, 5–18.

Yang, K. S. (1999). Towards an indigenous Chinese psychology: A selective review of methodological, theoretical, and empirical accomplishments. *Chinese Journal of Psychology, 41*, 181–211.

Yang, K. S. (2006). Indigenized conceptual and empirical analyses of selected Chinese psychological characteristics. *International Journal of Psychology, 41*, 298–303. doi:10.1080/00207590544000086.

Yu, C. F., & Huang, Y. T. (2009). Religious behavior of parents and parent-child relationships among adolescents from Christian families in Taipei city [父母宗教行為與青少年親子關係之研究-以台北市基督教家庭為例]. *Hwa Kang Journal of Agriculture, 23*, 35–60.

Yu, R. Y. (2011). Study of college students' mental characteristics [大学生精神性特点研究]. *Knowledge Economy,* 9, 152,156. doi:10.15880/j.cnki.zsjj.2011.09.095.