Distinguishing Between Self-Classified Religious and Spiritual Emerging Adults: Conceptual and Operational Challenges

By Paul Handal
Chelsi A. Creech†
Travis J. Pashak‡
Lea Caver†
Eunice Joy G. Perez
Michael G. Schwendeman
Christopher Griebel

The last ten years has witnessed a sharp increase in research related to spirituality as well as the proliferation of numerous measures of the construct. However, there has been little research examining the relationship between and among measures of spirituality, much less to measures of religion. This study sought to determine whether measures of spirituality and religion could discriminate between participants who self-identified as both religious and spiritual (B), religious only (R), spiritual only (S), or neither (N). Two religion measures and three spirituality measures assessed the constructs. It was predicted that those who self-identified as religious only would score significantly higher on the religion measures than those who identified as spiritual only, but those that identified as spiritual only would score significantly higher on the spirituality measures than those who identified as religious only. Results supported the first hypothesis, but not the second. The relationship between spirituality and religion measures suggested substantial overlap.

Introduction

Within the last decade of the 20th century, there was a significant decrease in church membership and attendance, which preceded an increase in spirituality (Hout & Fischer, 2002). Presently, a considerable percentage of

†Professor, Saint Louis University, USA.
‡Saint Louis University, USA.
§Saint Louis University, USA.
¶Saint Louis University, USA.
•Saint Louis University, USA.
Saint Louis University, USA.
Saint Louis University, USA.
people purportedly consider themselves spiritual but not religious—over a quarter of adults in the United States (Oman & Thoresen, 2007). It is possible that this pattern of decreased church membership and increased interest in spirituality is related to the continued secularization in the United States, which has separated spirituality from institutional religion. Within this paradigm of secularization, religion has come to be labeled as institutional, formal, objective, doctrinal, and authoritarian. Spirituality, on the other hand, represents an individual, subjective, and private relationship with a supreme being, higher power, or ultimate transcendence.

One result of the increase in interest in spirituality has been a sharp increase in research related to spirituality and its relationship to health, mental health, and adjustment. Despite the increased research pertaining to spirituality, the conceptualization and measurement of the construct remains a challenge (Kapuscinski & Masters, 2010). The conceptualization of spirituality has run the gamut from a construct that is separate and divergent from that of religion (Gorsuch, 1993) through a conceptualization that spirituality is broader than, and encompasses, religion (Kaiser, 2000), to the conceptualization that spirituality and religion are related constructs with overlapping characteristics (Zinnbauer, 1997), to yet another description that suggests that spirituality can never be separated from its traditional, religious roots (Koenig et al., 2012).

Given the differing conceptualizations of spirituality, it is not surprising that numerous scales have been developed to measure the construct. However, despite the existence of numerous scales, currently there is no agreed upon conceptualization or operational definition of spirituality, and there has not been much research conducted on the relationship between and among the measures of spirituality, much less their relationship to measures of religion.

This study was designed to address the relationship between and among measures of spirituality as well as to investigate the relationship between measures of spirituality and measures of religion. Additionally, this study was designed to investigate whether measures of spirituality and religion would discriminate between groups self classified as both religious and spiritual (B), religious only (R), spiritual only (S), and neither religious nor spiritual (N). This was accomplished by selecting three measures of spirituality and two measures of religion that were supported by appropriate reliability and validity data and that have been employed in previously published, peer-reviewed research.

It was hypothesized that the spiritual only group would score significantly higher than the religious only group on measures of spirituality and that the religious only would score significantly higher than the spiritual only group on measures of religion.
Method

Participants

Participants were 268 female undergraduate students from a private, religiously affiliated Midwestern university. The study included students between the ages of 18 and 24 (M=19.91, SD=3.47). Of the 268, 193 were Caucasian (72%). There were 145 Catholics (54%), 56 other Christians (21%), 24 were atheist or agnostic (9%), 11 were Hindus (4%), 5 were Muslims (2%), and 24 checked other religion (9%). The majority of participants described themselves as both spiritual and religious (N=159, 59%), while 39 participants described themselves as religious only (14%), 37 participants described themselves as spiritual only (14%), and 33 participants described themselves as neither spiritual or religious (12%).

Measures

Spirituality Measures

The Daily Spiritual Experiences Scale (DSES) (Underwood & Teresi, 2002) measured by a 16-item scale, which asks participants to indicate how often, they have certain spiritual experiences. Responses range from never (1) to many a day (6). Higher scores reflect a greater level of daily spiritual experiences (α = .96). Scores were summed for each respondent and then averaged across the 16 items. One of these items (“In general, how close do you feel to God?”) is reverse scored in a 4-point metric (not at all, somewhat close, very close, as close as possible) instead of a 6-point likert scale. To be consistent with the directionality (Underwood 2006), the raw score of this item is reversed coded and the 4-point scale is adjusted to fit the 6-point spectrum. The adjusted score is averaged for this subscale into the total for the resulting mean score. Further, the scale was divided into two subscales: a “theistic” subscale, with an alpha reliability of .95 and a “non-theistic (self-transcendent)” subscale, with an alpha reliability of .90 (Ellison & Fan 2008). Underwood and Teresi (2002) reported appropriate levels of validity for the measure, which were further supported by Ellison and Fan (2008) in their study on spirituality and well being.

The Spiritual Transcendence Scale (STS) is a 24-item scale, developed by Piedmont (1999), which consists of three subscales: universality, prayer fulfillment, and connectedness. Universality is the belief in the unity and purpose of life, prayer fulfillment is a feeling of joy and contentment that results from prayer or meditation, and connectedness is a sense of personal responsibility and connection to others. The scale items were answered on a 1 (strongly agree) to 5 (strongly disagree) likert-type scale. Piedmont (1999) showed these scales to have acceptable reliabilities: .83 for universality, .87 for prayer fulfillment, and .64 for connectedness.

The Spiritual Involvement and Beliefs Scale (SIBS) (Hatch, Burg, Naberhaus, & Hellmich, 1998) was designed to measure participants’ spiritual status. It consists of four sub-scales. The first is the external/ritual scale, which is a 13-item scale that reflect belief in a greater power. Second is the internal/fluid scale, with 11 items that reflect internal beliefs and growth. The
third is a seven item existential/meditative subscale that reflect existential issues. Finally is the humility/personal application subscale with four items that reflect humility and application of spiritual principles. Internal reliability statistics for three of the subscales are satisfactory: external/ritual, $\alpha=0.98$; internal/fluid, $\alpha=0.74$; existential/meditative, $\alpha=0.70$) but perhaps, as Hatch et al. (1998) suggest, not for humility/personal application sub-scale ($\alpha=0.51$). The internal consistency of the SIBS was reportedly high (Cronbach’s alpha = .92) and presented a test-retest reliability of $r = .92$. Validity was reported Hatch et al., (1998) via factor analysis and the relationship of the SIBS with the Spiritual Well-Being Scale.

**Religiousness Measures**

The PRI (Lipsmeyer 1984) is a 45-item, nine scale, multidimensional measure of religion. The scales measure personal prayer (PRP); ritual attendance (RA); non-ritual, church-related activity (NRA); belief in God (BLFGOD); belief in an afterlife (AFTLIFE); perceived congruence of a person’s religious beliefs with their attitudes on social and moral issues (RSM); the extent to which an individual's ideas about religion guide their philosophy or way of life (IDEO); the subjective experience of feeling close to God (CLOSEGOD); and integration or the extent to which persons perceive that their relationship with God influences their cognition , affect, and behavior (INT). Most of the items use a 6-point Likert response format; however, others use a multiple choice or yes/no format.

Lipsmeyer reported that test-retest reliability coefficients over a one-week period were between .83 and .97 for the nine scales in an adult population. Additionally, Lipsmeyer found that the PRI had high concurrent validity; religious professionals (e.g., priests, ministers, nuns) scored significantly higher on all scales than the general public. Also, Lipsmeyer reported that atheists, agnostics, and those with no religious preference scored significantly lower than other major religious groups. Lipsmeyer reported that each subscale of the PRI correlated highest with integration (INT), and that it had the highest stability coefficient and was the best single measure of religion. (Ross, Handal, Clark and Vander Wall 2009). Validity for the PRI is reflected by studies reporting its relationship to adjustment in adolescence (Mosher & Handal, 1997), emerging adults (Low & Handal, 1995), and adults (Crawford, Handal & Weaver 1989).

The Duke University Religion Index (DUREL) is a five-item measure of religious involvement that is incorporated in epidemiological surveys investigating the relationship between religion and health outcomes (Koenig & Bussing, 2010). This brief measure of religion was established for use in both cross-sectional and longitudinal studies. It evaluates three main dimensions of religion: organizational religious activity, non-organizational religious activity, and intrinsic/subjective religion. The scale assesses each of these components by a separate “subscale”, and correspondences between health outcomes should be examined by subscale in different models. The scale as a whole displayed high test-retest reliability (intra-class correlation = 0.91), high internal
consistence (Cronbach’s alpha’s = 0.78 – 0.91) and has high convergent validity with other religion measures.

**Demographic Measures**

The participants also completed a 22-item demographic questionnaire. These items asked about a participant’s age, ethnicity, sex, religious affiliation, college living arrangement, volunteer and work positions, and finally whether a participant identified as spiritual, religious or both, or none.

**Procedure**

Participants were recruited from undergraduate psychology classes. Some classes (approximately 66%) offered class credit for participation, while the other classes were not offered incentives for participation. Participants accessed the study via SONA, a university-approved research recruitment program, or through a link provided to them by professors who helped with recruitment. After accessing the study, they were directed to a link to the Qualtrics site that was hosting the survey. Participants first answered the demographic questionnaire. Next, the participants progressed through the Duke University Religious Index, the Personal Religion Inventory, the Daily Spiritual Experience Scale, the Spiritual Transcendence Scale, and the Spiritual Involvement and Beliefs Scale. Participants were encouraged to complete all sections in order, but were able to progress through the questionnaires at will. Participants were able to end the survey at any time, and were able to skip any questions they chose.

**Results**

**Table 1. Means and Standard Deviations for Each Variable and Significant Group Differences**

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>F Value</th>
<th>Significance</th>
<th>Order of means</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIBS</td>
<td>88.3918</td>
<td>14.75059</td>
<td>$F(3,13)=35.5$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R&lt;B</td>
</tr>
<tr>
<td>STS</td>
<td>85.3358</td>
<td>12.60250</td>
<td>$F(3,13)=22.8$</td>
<td>p&lt;.001</td>
<td>N&lt;S,R,B</td>
</tr>
<tr>
<td>DSES</td>
<td>58.1148</td>
<td>17.19940</td>
<td>$F(3,13)=27.6$</td>
<td>p&lt;.001</td>
<td>N&lt;S,R,B</td>
</tr>
<tr>
<td>PRP</td>
<td>16.1877</td>
<td>5.66589</td>
<td>$F(3,13)=27.4$</td>
<td>p&lt;.001</td>
<td>N&lt;S,R,B</td>
</tr>
<tr>
<td>RA</td>
<td>12.7870</td>
<td>5.37723</td>
<td>$F(3,13)=28.0$</td>
<td>p&lt;.001</td>
<td>N,S&lt;R,B</td>
</tr>
<tr>
<td>NRA</td>
<td>10.8195</td>
<td>4.54087</td>
<td>$F(3,13)=7.3$</td>
<td>p&lt;.001</td>
<td>N,S&lt;R,B</td>
</tr>
<tr>
<td>IDEO</td>
<td>8.2826</td>
<td>2.21233</td>
<td>$F(3,13)=12.8$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
<tr>
<td>RSM</td>
<td>31.4130</td>
<td>6.57790</td>
<td>$F(3,13)=19.2$</td>
<td>p&lt;.001</td>
<td>N,S,R,B</td>
</tr>
<tr>
<td>INT</td>
<td>74.6787</td>
<td>22.28508</td>
<td>$F(3,13)=22.0$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
<tr>
<td>BLF</td>
<td>5.1000</td>
<td>1.38271</td>
<td>$F(3,13)=42.6$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
<tr>
<td>AFT</td>
<td>4.1407</td>
<td>1.16748</td>
<td>$F(3,13)=31.1$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
<tr>
<td>CLS</td>
<td>3.1218</td>
<td>1.10047</td>
<td>$F(3,13)=34.3$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
<tr>
<td>DUREL3</td>
<td>10.0786</td>
<td>3.31623</td>
<td>$F(3,13)=29.7$</td>
<td>p&lt;.001</td>
<td>N&lt;S&lt;R,B</td>
</tr>
</tbody>
</table>

Note: N=neither spiritual or religious, S=spiritual only, R=religious only, B=both
Table 1 presents the means and standard deviations for each of the religion and spirituality measures. As can be seen from Table 1, the means and standard deviations of the spirituality measures are similar to those previously reported in the literature (Underwood 2006; Piedmont 1999; Hatch, et al. 1998). The means and standard deviations of the religion measures are also consistent with reported means and standard deviations (Lipsmeyer 1984; Koenig & Bussing 2010).

In order to determine the relationship between and among measures of religion and spirituality, Pearson correlations were computed. Table 2 reports these analyses.

<table>
<thead>
<tr>
<th>Table 2. Intercorrelations between the Spirituality and Religion Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>STSTOTAL</td>
</tr>
<tr>
<td>SIBSTOTAL</td>
</tr>
<tr>
<td>DSESTOTAL</td>
</tr>
<tr>
<td>PRP</td>
</tr>
<tr>
<td>RA</td>
</tr>
<tr>
<td>NRA</td>
</tr>
<tr>
<td>IDEO</td>
</tr>
<tr>
<td>RSM</td>
</tr>
<tr>
<td>INT</td>
</tr>
<tr>
<td>BLF</td>
</tr>
<tr>
<td>AFT</td>
</tr>
<tr>
<td>CLS</td>
</tr>
<tr>
<td>DUREL</td>
</tr>
</tbody>
</table>

** Correlation is significant at the 0.01 level (2-tailed).

As can be seen in Table 2, the three spirituality measures were highly correlated with each other, ranging from .79 to .80. In terms of religion measures, the DUREL is highly correlated with the single best measure of religion on the PRI, the INT (.78). Interestingly the DUREL is also highly correlated with each of the subscales on the PRI, with the exception of the NRA (.39) and RSM (.26) subscales. Surprisingly, the measures of spirituality and religion are highly correlated with one another. Using the INT subscale as the single best measure from the PRI, the spirituality measures are correlated at .71 (STS), .86 (SIBS) and .82 (DSES). Similarly the DUREL is highly correlated with the measures of spirituality .63 (STS), .76 (SIBS), and .77 (DSES).

In order to determine whether significant differences existed on measures of spirituality and religion between or among the self classified groups, a MANOVA was calculated; the independent variable was the four self-identified groups, namely, both spiritual and religious, religious only, spiritual only, and none. The results revealed a significant MANOVA $F(39, 806.20)=9.137$, $p<.001$. In order to determine which measures of spirituality and religion differed significantly, a series of ANOVA analyses were
computed using the same four self-identified groups as the independent variable and the spiritual and religious measures as the dependent variables. Results of these analyses are presented in Table 1 as well. As can be seen in Table 1, significant differences were found between the four self-identified groups on every spiritual and religious measure. In order to determine which self-identified groups differed from one another, a series of Tukey’s post hoc tests were computed. Results of these analyses are also in Table 1.

For the spirituality measures the results revealed that only the SIBS significantly differentiated between each of the four groups, as seen in the last column of Table 1. However, contrary to expectation, the highest scoring group on the SIBS measure was the both spiritual and religious group, followed by religious only, followed by spiritual only, and finally the neither group. For the STS, the both spiritual and religious group was significantly higher than both the religious only and the spiritual only groups, which did not differ from each other, and the neither group significantly differed from all three groups. The final spirituality measure, the DSES, revealed that there were no significant differences between the both spiritual and religious group and the religious only group, but both of these groups differed significantly from the spiritual only group and the neither group, which also differed from each other.

In terms of the measures of religion, there is a clear pattern of results in terms of which groups differ from one another. On each measure of religion, there is no significant difference between the both spiritual and religious group and the religious only group. Also, those two groups differ significantly from the spiritual only group and the neither group on every measure of religion, with the exception of RSM, on which there are no significant difference between or among the four groups. Within this pattern, it is common that the both religious and spiritual group and the religious only group score significantly higher than the neither group with two exceptions. On ritual attendance and non-ritual attendance, the spiritual only and neither group do not differ significantly from each other. Otherwise, the pattern is that the both spiritual and religious group and the religious only group score highest on each measure of religion and typically differ significantly from the spiritual only and neither groups.

Discussion

The results from this study partially supported the original hypotheses, particularly in relation to the religion measures. Specifically, there was a clear pattern in the results, with the B group and the R group scoring highest, significantly higher than both the S and N group. The S group also scored significantly higher than the N group. This pattern provides clear support for the validity of the religion measures.

However, the results for the spirituality measures did not offer support for the hypotheses. The results from the spirituality measures presented the same
clear pattern as the religious measures, with the B and R groups scoring the highest, followed by the S and N groups. These results do not support the validity of these instruments as measures of spirituality. Support would have occurred if the spiritual only group had scored significantly higher than the religious only and neither groups. Rather, it appears that the spiritual measures are not significantly different than the religious measures. This observation is supported by the correlations between the religion and spirituality measures. The spirituality measures correlate with each other from .74 to .80. However, the spirituality measures also correlate with the religious subscale of Integration (which is the best single subscale measure of religion on the PRI) from .71 to .86, and with the DUREL from .63 to .77. These correlations indicate a significant amount of common variance. The correlations reported indicate that at a minimum, there is 40% shared variance, and possibly as much as 74% shared variance between the religion and spiritual instruments. The issue of validity of spirituality measures is not a new one, having been most recently addressed by Kapuscinski and Masters (2010), who discussed conceptual problems, scale development, and sample selection problems with measures of spirituality.

In terms of conceptualization, Kapuscinski and Masters (2010) observed the necessity to clearly conceptualize the concept of spirituality, which has been done in two very different manners. The first entails conceptualizing spirituality as distinct from religion and not necessarily including a supreme or divine being, but rather conceptualizing spirituality as a transcendence from oneself (Astrow, Pulchalski, & Sulmasy 2001). The second entails conceptualizing spirituality as overlapping with the construct of religion and including a supreme or divine being (Kaiser 2000). The latter conceptualization of spirituality and its relationship to religion appears to be parallel and consistent with the history of both concepts. Historically, the three major world religions of Judaism, Islam, and Christianity have each conceptualized religion and spirituality as two highly integrated constructs. Religion, i.e., the institutional, creational, ritual, was always present as a means from which to develop and maintain a closer, personal relationship with the divine (i.e., spirituality). Perhaps, as the North American culture became more secular, in terms of separating religion and spirituality, psychology followed the culture and created two separate constructs that may have historically been overlapping (Donahue 1997). If that is the case, our results are not surprising, since they seem to suggest that in defining religion as objective, institutional, and creational, we defined half of what religion is, since it has always included a relationship with a higher power, closeness to a deity, a subjective nature, and personal prayer. For example, within Catholicism, there are many different spiritualities, one of which is Ignatian Spirituality, which presents a method of prayer, through which one can attain a close, personal relationship with Jesus.

It appears that given the overlap between our measures of religion and spirituality, a salient question exists as to whether the measures of spirituality add significantly to the prediction of mental health, health, and adjustment after the contribution of religion has been determined. This may be an appropriate...
next step for future research in this area. Future research may also include an analysis of which, if any, items on the spirituality measures discriminate between religious only and spiritual only groups. This could result in a more discriminating measure of spirituality. Another way to address the issue of measurement is to engage in qualitative research with individuals who self identify as religious only and spiritual only in order to better understand the ideographic components of their definition and then to determine commonalities. 

In terms of the measures of religion, the one subscale that did not differentiate between the four groups was the RSM on the PRI. It is highly likely that this measure is neither a measure of religion or spirituality. This observation would be supported by the correlations between the RSM subscale and other measures of religion (i.e., the DUREL, where the correlation is .42 and as low as .25 with the Belief subscale and .39 with the Integration of the PRI) and the three measures of spirituality where the correlations are .33 (STS), .38 (SIBS), and .39 (DSES).

Interestingly, the results pertaining to the neither group were predictable in that the neither group typically scored significantly lower than the other groups. It is possible that this finding is accounted for primarily by the presence of religion in both the spiritual and religious instruments. In reviewing the items that comprised the three spirituality measures, it is apparent that they contain items that would be consistent with current operational definitions of religion, referring to a deity, an afterlife, or worship services. For example, the STS has a number of items that tap belief in afterlife and prayer—not necessarily ritual or personal prayer—and on this measure, the religious participants did not differ from the spiritual participants. The DSES also has these types of items, as well as asking about worship services, which is a religious item rather than spiritual. The SIBS, while not asking directly about religion, has a number of items that are tied to creedal religious beliefs, e.g., “I can find meaning in times of hardship,” “My life has a purpose,” “I believe there is a power greater than myself,” making it harder to discriminate between religion and spirituality.

While our results present a clearly defined pattern between and among the four self identified groups on measures of religion and spirituality, these results were found in a sample of emerging adult females enrolled in a private, religiously affiliated university. Consequently, the issue of generalization becomes salient. Future research should be conducted to determine whether these results apply to a more diverse population of emerging adults—namely those enrolled in public universities, community colleges, and those female emerging adults not enrolled in higher education—as well as older women. Additionally, it unclear whether the pattern of results reported here would be found in an emerging adult male sample.
References


