PSYCHOMETRIC PROPERTIES OF THE DUKE UNIVERSITY
RELIGION INDEX, ENGLISH AND SPANISH VERSIONS, FOR
HISPANIC-AMERICAN WOMEN

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology

by
Patricia Wansley Taylor
Summer 2013
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Patricia Wansley Taylor:

Psychometric Properties of the Duke University Religion Index, English and
Spanish Versions, For Hispanic-American Women

Vanessa Malcarne
Vanessa Malcarne, Chair
Department of Psychology

Scott Roesch
Department of Psychology

Elva Arredondo
Graduate School of Public Health

June 28, 2013
Approval Date
Copyright © 2013
by
Patricia Wansley Taylor
All Rights Reserved
ABSTRACT OF THE THESIS

Psychometric Properties of the Duke University Religion Index, English and Spanish Versions, For Hispanic-American Women
by
Patricia Wansley Taylor
Master of Arts in Psychology
San Diego State University, 2013

The central aim of the present study was to investigate the psychometric properties of scores on the DUREL in its English version and Spanish translation for Hispanic-American (HA) women. The Duke University Religion Index (DUREL) seeks to measure religiosity and offers a total score of religiosity and three subscales scores for organizational religiosity, non-organizational religiosity, and intrinsic religiosity. A community sample of 488 self-identified HA women completed the DUREL in English (n = 245) or Spanish (n = 243), per each woman’s expressed language preference. Normative data for DUREL total and subscale scores were described for more acculturated (English preference) and less acculturated (Spanish preference) HA women, and for the total group. Cronbach’s alpha suggested good internal consistency of the total scale for the English-preference and total group samples, but data from the Spanish group was less internally consistent. DUREL subscales of organizational and non-organizational religiosity have only one item each and, therefore, could not be assessed for internal consistency, but the three-item subscale of intrinsic religiosity had satisfactory internal consistency for the total sample and each of the two language-preference subgroups. Exploratory factor analysis (EFA), found support in the sample and in both subgroups for a model of one umbrella factor of religiosity with three underlying factors of organizational religiosity, non-organizational religiosity and intrinsic religiosity. For the total sample and both subgroups, EFA suggested one factor for the items of the intrinsic religiosity subscale. Confirmatory factor analysis validated our expectation for an overarching one-factor structure for the total sample and the English-preference subgroup. However, the one-factor structure was not confirmed for the Spanish-preference subgroup, disputing cross-language measure invariance and challenging the validity of the DUREL. Age was positively related to total score, and all subscale scores of religiosity, for the total sample and the English-preference subgroup. For the Spanish preference sample, age was positively related to non-organizational religiosity, but did not relate significantly to organizational religiosity or intrinsic religiosity. These findings suggests that the response pattern of less acculturated HA women to the DUREL’s Spanish version is different from the response pattern of more acculturated HA women to the DUREL’s English version.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>ix</td>
</tr>
<tr>
<td>REVIEW OF THE LITERATURE</td>
<td>1</td>
</tr>
<tr>
<td>Religiosity as an Important Psychosocial Health Variable</td>
<td>2</td>
</tr>
<tr>
<td>Good Measures are Needed</td>
<td>3</td>
</tr>
<tr>
<td>The DUREL</td>
<td>4</td>
</tr>
<tr>
<td>History of the DUREL</td>
<td>5</td>
</tr>
<tr>
<td>Description and History of DUREL Subscale of Organizational Religiosity: Item One</td>
<td>6</td>
</tr>
<tr>
<td>Description and History of DUREL Subscale of Non-Organizational Religiosity: Item Two</td>
<td>6</td>
</tr>
<tr>
<td>Description and History of DUREL Subscale of Intrinsic Religiosity: Items Three, Four and Five</td>
<td>7</td>
</tr>
<tr>
<td>Scoring the DUREL</td>
<td>8</td>
</tr>
<tr>
<td>Psychometric Properties of the DUREL (English Version)</td>
<td>8</td>
</tr>
<tr>
<td>Reliability</td>
<td>8</td>
</tr>
<tr>
<td>Validity</td>
<td>9</td>
</tr>
<tr>
<td>Sociodemographic and Health Influences</td>
<td>9</td>
</tr>
<tr>
<td>Limitations of the DUREL</td>
<td>10</td>
</tr>
<tr>
<td>Psychometric Properties of the DUREL in Other Ethnic Population Groups, Including Hispanic-Americans</td>
<td>11</td>
</tr>
<tr>
<td>Importance of Availability of Valid Spanish Measures in Health Research</td>
<td>12</td>
</tr>
<tr>
<td>Purpose of the Present Study</td>
<td>13</td>
</tr>
<tr>
<td>METHODS</td>
<td>14</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td>Measures</td>
<td>14</td>
</tr>
<tr>
<td>Procedures</td>
<td>14</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Duke University Religion Index, English and Spanish Versions ..................................15
Table 2. Item Response Descriptives for Sample of Hispanic Women and English and Spanish Subgroups ..................................................................................................................25
Table 3. Subscale Descriptives for Sample of Hispanic Women and English and Spanish Subgroups ..........................................................................................................................26
LIST OF FIGURES

PAGE

Figure 1. One-factor model of religiosity for DUREL scale. ..................................................19
Figure 2. Three-factor model of religiosity for DUREL scale.................................................19
REVIEW OF THE LITERATURE

A mounting body of evidence indicates that the psychosocial examination of religiosity is important to our understanding of health outcomes for the individual and the community (John E. Fetzer Institute & National Institute on Aging, 1999; Padilla & Villalobos, 2007). Until recently, research has been hindered by the lack of measures of religiosity that were brief, methodologically sound, and easily used in research or clinical settings, while at the same time capturing the major elements identified as important to research on health and aging (Klemmack et al., 2007). The Duke University Religion Index (DUREL, also known as the DRI; Koenig, Parkerson, & Meador, 1997) was created to help fill this gap. The DUREL quickly and comprehensively assesses a person’s religious involvement (Storch, Roberti, Heidgerken, et al., 2004).

Widespread interest in the link between health and religion or spirituality is demonstrated by numerous publications in both scholarly journals and general interest magazines (John E. Fetzer Institute & National Institute on Aging, 1999). This interest takes place amid a culture of dynamic and thriving religious beliefs and practices. According to recent Gallup polls, forty percent of Americans are classified as very religious, based on their responses that “religion is an important part of their daily life and that they attend religious services every week or almost every week” (Newport, 2013). Another 29% are considered moderately religious, responding that “religion is important in their lives, but that they do not attend services regularly or that religion is not important but that they still attend services” (Newport, 2013). Illustrating the fluid nature of U.S. religious affiliations, twenty-eight percent of U.S. adults have moved from the religion in which they were raised to another one or to practicing no religion (Pew Research Center, 2008). The percentage reporting that they are currently unaffiliated with a particular faith (16.1%) is more than double those who report they were unaffiliated as children. However, a substantial percentage (almost 4% of adults) also state they were unaffiliated as children but have come to identify with a religious group.

Religious affiliation or lack of affiliation, however, does not capture the religiosity of a population. Religiosity incorporates cognitions and behaviors within personal, social and organizational settings and uniquely can be characterized both personally and culturally.
Due to its multidimensional nature, the study of religiosity is engaging, complex, and fertile for scientific investigation.

**RELIGIOSITY AS AN IMPORTANT PSYCHOSOCIAL HEALTH VARIABLE**

Since the late 1980s, empirical studies dedicated to the connection between religiosity or spirituality and health began to rise (Lee, Zahn, & Baumann, 2011). By 2001, twelve hundred studies had quantitatively examined some aspect of the relationship between health and religion. In the subsequent ten years, two thousand more studies were published, even though this area of research remains young, somewhat marginal, and poorly funded (Koenig & Büssing, 2010). The majority of this research comes from English speaking countries (Lee et al., 2011).

Most researchers find positive relationships between psychological and physical health (e.g., Arredondo, Elder, Ayala, Campbell, & Baquero, 2005). In fact, numerous studies have found a significant, but weak, positive association between physical and mental health and religious or spiritual contents (Koenig, 1998; Lee et al., 2011). In other words, greater scores of religiosity tend to correlate with better mental and physical health. For many patients, religion and spirituality hold coping reservoirs that help them manage life stressors, including their illnesses (Verghese, 2008).

Research design, methods, and outcome considered may explain the inconsequential link between religious involvement and subjective physical health found in a few studies (Arredondo et al., 2005). Early research with poor methodology contributed to formerly negative views of religion’s impact on mental and physical health. Views have changed as research using large random or systematic samples of mentally stable adults increased (Koenig & Larson, 2001). However, lack of adequate assessment measures has delayed the process and progress of research. Changes in medical practice (Koenig & Larson, 2001; Verghese, 2008) have also urged a practical tool to assess religiosity.

Successful healthcare strategies invite professionals to integrate cultural factors into health promotion activities (Padilla & Villalobos, 2007). Both practitioners and researchers should be collecting valuable psychosocial information that includes the cultural aspects of individuals, including their level of acculturation and religiosity (Padilla & Villalobos, 2007).
Spirituality is one issue Mexican-Americans report has an effect on the quality of their medical encounters and preventative behaviors. In the practice of health care in the Mexican-American culture, *curanderismo* (folk-healing practices that often include spiritual and emotional elements) continues to be important. *Curanderismo* is a good example of how Mexican-Americans regularly blend cultural strategies with conventional Western medicine. Many Mexican-Americans do not view spirituality as incompatible with modern medicine (Padilla & Villalobos, 2007).

**GOOD MEASURES ARE NEEDED**

In 1986, researchers began to call for more research on the link between mental health and religion (Koenig et al., 1997). Though research on the connection between mental health and religion was on the rise, well-designed and theoretically grounded studies were infrequent (Koenig et al., 1997). As suggested by Koenig and colleagues (1997), one contributory factor was researchers’ uncertainty on how to measure religiousness comprehensively, quickly, and inoffensively. To advance, the research required validated measures that were brief, theoretically coherent, and suitable for a variety of populations in healthcare, including very ill or weak patients (Sherman et al., 2000). The increased research interest in and evidence of the relationships between religion and health (Hays, Meador, Branch, & George, 2001; John E. Fetzer Institute & National Institute on Aging, 1999; Klemmack et al., 2007; Koenig & Larson, 2001; Koenig et al., 1997; Levin, 1994; Sherman et al., 2000; Sherman et al., 2001; Storch, Roberti, Bravata, & Storch, 2004; Storch, Roberti, Heidgerken, et al., 2004), and subsequent recommendations in medical practice, to assess and attend to patients’ religiousness/spirituality (Klemmack et al., 2007; Verghese, 2008) moved the need for a measure such as the DUREL from the theoretical and unhurried to the practical and exigent.

In 1995, the National Institute on Aging (NIA) and the Fetzer Institute brought together an expert panel of researchers in religiousness/spirituality, and health/well-being for a conference (John E. Fetzer Institute & National Institute on Aging, 1999). The group’s purpose was to address the complete lack of available standardized measures of religiosity that experts could recommend to health researchers, while evidence demonstrating links between health outcomes and spiritual or religious variables continued to grow. The goal, as
encouraged by The National Institute on Aging, was to develop rigorous, parsimonious scales and indices for use in studies of health and aging (John E. Fetzer Institute & National Institute on Aging, 1999).

Conceptualization and measurement, however, were lingering thorny issues (John E. Fetzer Institute & National Institute on Aging, 1999; Sherman et al., 2001). Social and behavioral scientists generally view religiousness as a dynamic and multidimensional concept (Klemmack et al., 2007; Storch, Roberti, Heidgerken, et al., 2004), involving beliefs, practices and personal devotion (Klemmack et al., 2007). However, at the time, few health researchers had a research background in religiousness/spirituality (John E. Fetzer Institute & National Institute on Aging, 1999). Most health researchers were unfamiliar with historical efforts to conceptualize and measure the multidimensional nature of religiousness.

As an additional challenge and despite its multi-dimensional character, religion historically had been assessed as only one variable, usually religious denominational affiliation (Klemmack et al., 2007; Sherman et al., 2001; Storch, Roberti, Heidgerken, et al., 2004), with researchers occasionally also assessing church attendance (Sherman et al., 2001). Unfortunately, denominational affiliation may more accurately reflect heritage or parental affiliation than personal beliefs, practices and devotion, particularly for younger or less religious persons. The Fetzer Institute/National Institute on Aging working group agreed that the variable “religiosity” could not be measured that simply. Instead, they recommended examining relevant dimensions of religiousness and spirituality separately for their effects on health (John E. Fetzer Institute &

**THE DUREL**

The creation of the DUREL (Koenig et al., 1997) followed years of psychometric problems with previous measures (Sherman et al., 2000; Sherman et al., 2001; Storch, Roberti, Heidgerken, et al., 2004). Its arrival filled a need for a brief, easy-to-use, non-intrusive, multi-dimensional measure of religiosity (Klemmack et al., 2007; Sherman et al., 2000; Sherman et al., 2001; Storch, Roberti, Heidgerken, et al., 2004). Moreover, thus far, the DUREL has also proven to be a valid and theoretically sound measure that can be easily embedded within other more general research in health or aging (John E. Fetzer Institute &
National Institute on Aging, 1999; Sherman et al., 2000; Sherman et al., 2001; Storch, Roberti, Heidgerken, et al., 2004).

Although experts have had disagreements concerning particular dimensions of religiousness/spirituality, and which dimensions are most germane to research on physical and mental health and linked to health outcomes, three dimensions are always identified: organizational involvement, nonorganizational involvement, and intrinsic religiousness (Klemmack et al., 2007; Koenig & Büssing, 2010). These correspond perfectly with the three subscales of the five-item DUREL, as identified at the conference by the National Institute on Aging and Fetzer Institute conference (Koenig & Büssing, 2010). With only five self-report questions, the DUREL was designed to define and capture these three prominent dimensions of religious involvement (Koenig et al., 1997).

The measure provides a score for each subscale as well as providing a score for overall religious involvement (Koenig & Büssing, 2010; Koenig et al., 1997; though see discussion below regarding new recommendations against using a total score). Although the DUREL does not address other important religious dimensions, such as religious coping, daily spiritual experiences, or spirituality, it fills the gap for a practical measure of religiosity while capturing the components most researchers acknowledge as central to any study of religiousness (Klemmack et al., 2007).

**History of the DUREL**

Items One and Two of the DUREL stem from large community and clinical studies supported by the National Institutes of Health (Koenig & Büssing, 2010; Koenig et al., 1997). For OR (Item One), see also Allport and Ross (1967), Genia (1993), Hoge (1972), and Peña and Frehill (1998). For NOR (Item Two), see also Allport and Ross (1967), Hoge (1972), and Levin (1999). The three items creating the third subscale IR come from Dean Hoge's (1972) 10-item Intrinsic Religiosity Scale (Koenig & Büssing, 2010). Hoge’s scale in turn was derived from Allport’s and Ross’s (1967) twenty-item Religious Orientation Scale (ROS). Developers of the DUREL administered Hoge’s scale to 455 medical inpatients. Using regression, they examined the relationship between each scale item and five health factors (severity of mental illness, functional status, depressive symptoms, speed of recovery from depression, and social support). Using principal components factor analysis, two major
factors were revealed (one intrinsic and one extrinsic). They then chose three items from the Hoge scale, based on their loadings on the intrinsic factor, relationship with total score, and correlations with health outcomes. These three items were added to the existing two single-item subscales to form the final five-item index (Koenig & Büssing, 2010; see also Allport & Ross, 1967; Brewczynski & MacDonald, 2006; Genia, 1993; Hoge, 1972). In 1997, the *American Journal of Psychiatry* published the DUREL (Koenig et al., 1997).

**Description and History of DUREL Subscale of Organizational Religiosity: Item One**

The first item of the DUREL measures “organizational religiosity” (OR) and can also be described as frequency of attendance at religious meetings (Koenig et al., 1997). OR refers to a person’s involvement with a formal religious organization, such as a church, synagogue, ward, mosque or ashram. It is a well-established measure of religiosity that includes both attitudinal and behavioral dimensions. OR may also be referred to in the literature as public religious activities, participation in formal religious institutions, and social religiosity. The DUREL measures OR by one question: “How often do you attend church or other religious meetings?” The response options for attendance frequency are presented in a frequency scale from 1 = more than once per week to 6 = never. The scale is reversed in the Spanish version, so that 1 = never and 6 = once per week. Item One comprises the first subscale (OR) of the DUREL.

**Description and History of DUREL Subscale of Non-Organizational Religiosity: Item Two**

The second item measures “non-organizational religiosity” (NOR) (Koenig et al., 1997), which could also be described as frequency of private religious and spiritual activity, such as prayer, meditation, or reading the Bible. NOR examines private religious and spiritual practices. Though these behaviors may occur in a formal institutional setting, in contrast to organizational religious involvement, they do not require organized religion, fixed times, fixed locations or fixed liturgical formulae. The focus is on private behaviors that occur individually or within a family, outside the context of collective experiences within organized religion (Krause, 1993; Levin, 1999). This domain is also referred to in the literature as non-organizational religious involvement, non-organizational religious activity,
private religious activity, private religious involvement, private religious practices, informal religiosity, and non-institutional religiosity. The DUREL assesses NOR with one question: “How often do you spend time in private religious activities, such as prayer, meditation or Bible study?” On both English and Spanish versions, this question is measured on a reverse-coded frequency scale ranging from 1 = more than once a day to 6 = never. Item Two makes up the second subscale (NOR) of the DUREL.

**Description and History of DUREL Subscale of Intrinsic Religiosity: Items Three, Four and Five**

Items Three, Four and Five measure “intrinsic religiosity” (IR; Koenig et al., 1997), which refers to the pervasiveness of spiritual/religious influence in daily life and decisions. Most researchers agree that IR is multidimensional (Williams, 1999). IR is construed as the extent to which a person incorporates his religious beliefs within his life or uses them to guide life decisions (Storch, Roberti, Heidgerken, et al., 2004; Storch, Strawser, & Storch, 2004). Items of IR are “designed to measure the importance of and commitment to one’s religious/spiritual beliefs” and “a general orientation to all aspects of life and social relationships” (Williams 1999). This domain is considered either as synonymous with religious commitment or as one of the elements of religious commitment (Leak & Fish, 1999; Williams 1999). Measuring IR provides a distinction between other domains of religiosity (including OR and NOR) that may be more heavily influenced by custom, culture and social setting, thus allowing further comparison and finer distinction between individuals within the same religious group or those performing the same religious behaviors. Items Three, Four and Five create the third and final subscale of the DUREL. In contrast to the DUREL’s first two item, which are questions, Items Three through Five are formatted as statements ((3) In my life, I experience the presence of the Divine (i.e., God); (4) My religious beliefs are what really lie behind my whole approach to life; and, (5) I try hard to carry my religion over into all other dealings in life.). Each item is assessed by means of a five-point frequency scale (1 = Definitely true of me; 2 = Tends to be true; 3 = Unsure; 4 = Tends not to be true; and 5 = Definitely not true).
Scoring the DUREL

Scoring the DUREL is very important for analysis and for interpretation. Originally, the developers of the DUREL advised creating a total score that ranged from five to 27, in addition to considering the three subscale scores (Koenig et al., 1997). However, in a recent article, the developers of the DUREL are now advising against creating a total summary score by adding up the three subscales. Instead, they advise examination of each subscale score independently in separate regression models to examine relationships to health outcomes (Koenig & Büssing, 2010). They argue that, if subscales are taken as a single model, multiple collinearity between them may cause a misinterpretation of the effects accompanying each subscale. Moreover, the subscales could cancel the effects of one another if combined into a single score or analysis (Koenig & Büssing, 2010). In 2000, Sherman and colleagues similarly recommended that the differing relationships of the three subscales of the DUREL with health outcomes justify examining the subscores separately rather than examining only total scores. However, they did not recommend against creating a summary score. Koenig and Büssing’s (2010) current recommendation is a departure from the developers’ previous argument (Koenig et al., 1997) that the subscales, although representing different subconstructs, can be considered together as an index of religiosity.

Psychometric Properties of the DUREL (English Version).

Previous study of the DUREL’s English version included examination of reliability, validity, influences from sociodemographics and health status, and some limitations of the DUREL.

Reliability

Although the dynamic nature of religiosity is an important dimension in the overall study of the psychology of religiosity and in the study of health, for application to oncology care, for example, short-term temporal stability is of primary importance (Sherman et al., 2001). A two-week test-retest in a small ($n = 20$) college undergraduate sample indicated high short-term reliability (0.91 intra-class correlation coefficient) (Storch, Strawser, et al., 2004). Cronbach’s alphas for the total scale fell between .78 and .92, indicating high internal
consistency (Koenig & Büssing, 2010; Koenig et al., 1997; Sherman et al., 2000; Sherman et al., 2001).

**Validity**

Support for a single factor structure was found by Storch, Roberti, Heidgerken, et al. (2004) using confirmatory factor analysis with two samples, indicating construct validity for one factor (“religiosity”). They did not find support for a two or three factor model.

Researchers have found high convergent validity with other measures of religiosity ($r = 0.71–0.86$), including the Intrinsic Religious Motivation Scale (IRMS), Age Universal Religious Orientation Scale (AUROS) - Intrinsic, Santa Clara Strength of Religious Faith (SCSRF), and Santa Clara Strength of Faith – Short Form (SCSRF – SF), with correlations between .78 to .86 (using Spearman correlations for all but SCSRF-SF which used Pearson) (Koenig & Büssing, 2010; Sherman et al., 2000; Storch, Roberti, Bravata, et al., 2004). However, in a sample of gynecology patients (women), the DUREL had only a modest (.28) relationship with the AUROS-Intrinsic (Sherman et al., 2000). Relationships with the AUROS-Extrinsic have been the smallest (-.10 to -.14) (Sherman et al., 2000).

Discriminant validity has also been reviewed. Sherman and colleagues (2000) examined correlations among the subscales of the DUREL, and found moderate correlations, ranging from .47 to .67, which suggested significant overlap among the scales. They found the DUREL to show no association with measures of quality of intimate relationships (using the Dyadic Adjustment Scale-Cohesion Subscale) or negative affect (from the Life Orientation Test), both predicted as conceptually unrelated to religiosity. Scores for social desirability (using the Marlowe Crowne Social Desirability Scale) correlated significantly with DUREL total and IR scores of cancer patients, but no significant association was found between social desirability scores and DUREL total or IR scores with gynecology patients. Correlations with social desirability were not found for DUREL OR or NOR subscales for either the gynecology or oncology groups.

**Sociodemographic and Health Influences**

The English DUREL’s scores have not shown effects from sociodemographic influences, including religious affiliation (Catholic or Protestant), marital status, or
employment status in samples of gynecological patients, or by sex in oncological patients (Sherman et al., 2000). Also DUREL total score and NOR score were unaffected by health influences, such as stage at diagnosis, elapsed time from diagnosis, status as in-patient or outpatient, previous number of transplants, or treatment phase (Sherman et al., 2000).

**Limitations of the DUREL**

The DUREL has several important design flaws. Two of the three subscales use only one item each. Using one item as a subscale causes problems in model analysis and influences accuracy of capturing the target concept. Confirmatory factor analysis (CFA) cannot be used to test DUREL scores for fit to the theoretical three-factor model because more than one item per subscale is required for the CFA’s internal correlations. Additionally, a three-factor model cannot be tested with CFA using scores for the three subscales because that model is just-identified and, therefore, meaningless.

As discussed in Schwartz (1999), the unstandardized response frequency scales between items and subscales likely adds variance to the data collected data. Furthermore, it requires extra steps to standardize the response scales before most analyses. Other than maintaining the same historical response frequency scales that came with the items, there seems to be no useful reason for constructing the measure with these variations that accompanied the items’ historical origins.

The measure is heavily weighted in favor of IR. First, the IR subscale consumes three of the five questions of the DUREL. Then, due to the variances in frequency, response scales, (one to five or one to six, as discussed above), each item has more potential points per question. In addition, the debate on the definition and interpretation of intrinsic religiosity remains ongoing (Koenig & Büssing, 2010), which may influence perception of the third subscale of the DUREL.

Item Two of the DUREL asks, “How often do you spend time in private religious activities, such as prayer, meditation or Bible study?” The wording of this item does not account for varied meanings in current usage today. Meditation is commonly practiced, prescribed and taught as a secular activity, such as to practice and increase concentration and mindfulness. Additionally, many churches today use the phrase “Bible study” to designate a organized meeting of parishioners, who might meet as a small group in a home or church building, for
example, or as a very large group in a meeting virtually indistinguishable from a “Sunday sermon.” Due to the wording in the DUREL’s Item Two that does not reflect current usage, responses may be unintentionally capturing both organizational religiosity and secular behavior.

The DUREL was designed for measurement of religious participation in Western religions, such as Christianity, Judaism and Islam, which is also a limitation (Koenig & Büsing, 2010). Accuracy may decrease in assessment of persons with other religious traditions, such as those practicing Buddhism, Hinduism, individualized secular humanistic spirituality, pantheism, or espirito, for example. For these populations, the DUREL may still be valid, but the wording may need to change to accommodate the cultural traditions, e.g., replacing the word “Bible” by “writings of Buddha” or “Tao Te Ching”. Continued research on varied populations is needed to determine this question, and whether the DUREL can be used to sufficiently represent diverse spiritual practices and orientations.

The DUREL affords a brief assessment of religiosity. For a more in-depth assessment, other measures are needed (Koenig & Büsing, 2010). In addition, the measure does not distinguish or allow for identification of unhealthy religious behaviors, such as such frequent participation in an organization that it creates distress and disconnection from social and familial ties, or participating in religious practices identified as detrimental to health. These would seem to be important in a measure used with health outcomes and for use with diverse populations.

**Psychometric Properties of the DUREL in Other Ethnic Population Groups, Including Hispanic-Americans.**

To date, psychometric properties of the English version of the DUREL have not been examined in diverse groups. The literature calls for further exploration of the DUREL’s psychometric properties in more heterogeneous populations (i.e., beyond American, primarily Caucasian, middle-class, Protestant or Catholic persons on whom the DUREL’s English version primarily has been validated) (Sherman et al., 2000; Storch, Roberti, Heidgerken, et al., 2004). Previous research indicates that differences in culture, ethnicity, gender, income, age and religious affiliation may affect scores of total religiosity as well as religiosity subscales (Arredondo et al., 2005; Ghorpade, Lackritz, & Singh, 2006; Sherman et al., 2001). For example, less acculturated Mexican-American (and American Indian) samples
had more intense religious experiences than did Caucasian or acculturated Mexican-Americans (Hood & Hall, 1977). Among Blacks with cancer, less depressive symptomology was linked with religious activity. In addition, effects of participation in religious activity were stronger for Blacks than Whites (Musick, Koenig, Hays, & Cohen, 1998). Additionally, IR has been negatively correlated with acculturation, meaning that less acculturated persons tended to have higher levels of IR (Ghorpade et al., 2006).

The DUREL has been translated into 12 languages (Koenig & Büssing, 2010), including Spanish, Chinese, Romanian, Japanese, Thai, German, Norwegian, Dutch, Danish, Malay (Nurasikin, Aini, Aida Syarinaz, & Ng, 2010), Portuguese (Carvalho, Taunay, & Gondim, 2012; Lucchetti et al., 2012), and Farsi/Persian (Hafizi et al., 2013; Saffari, Zeidi, Pakpour, & Koenig, 2013). Support for a unidimensional construct of the DUREL in other translations was found for the Malay translation (Nurasikin et al., 2010), and for the Farsi translation (Hafizi et al., 2013; Saffari et al., 2013). Although researchers are using a Spanish translation of the DUREL, no psychometric data have been published on the validity of the translation.

**IMPORTANCE OF AVAILABILITY OF VALID SPANISH MEASURES IN HEALTH RESEARCH**

The importance of valid Spanish language measures is increasing, particularly for researchers and clinicians in the United States (e.g., Ramirez, Ford, Stewart, & Teresi, 2005). The U.S. holds the largest Spanish-speaking community outside Mexico (Mantilla, 2008). Thirty-seven million persons (12.8% of U.S. population) aged five or older speak Spanish (or Spanish Creole) as the primarily language in the home (United States Census Bureau, 2010), a figure that has doubled since 1990. It is the second most used language in the U.S. About half of these residents (55.3%) report that they speak English very well.

The creation of valid, suitable measures, and measurement accuracy for this population, are important for health research (Ramirez et al., 2005). Language is an essential part of patient communication within the healthcare system than can help to form a bridge to healthcare for the often underserved population of Spanish speakers (Padilla & Villalobos, 2007). Increasing our understanding of factors influencing healthcare in this population, by measures such as the DUREL, provides researchers and clinicians with better tools to serve this population and adds to the scientific body of knowledge.
PURPOSE OF THE PRESENT STUDY

The central aim of the present study is to describe the psychometric properties of the original English version and the Spanish translation of the DUREL for Hispanic-American (HA) women, and thus to determine its validity for use with HA women. The specific objectives of our examination were as follows. First, to provide normative data for the DUREL in more acculturated (English-language preference) HA women, as well as less acculturated (Spanish-language preference) HA women, we described scores, including subscale scores and item scores for the total sample (HA-T) and the two language subgroups (HA-E and HA-S, for English and Spanish preference subgroups, respectively; Objective 1). Reliability (Objective 2) and structural validity (Objective 3) were examined for the total sample and both subgroups. Then, we compared and evaluated psychometric properties of the English and Spanish versions of the DUREL for cross-language measure invariance (Objective 4). Finally, we examined relationships between religiosity scores and age for the three groups (Objective 5).
METHODS

Our research approach is outlined as follows.

PARTICIPANTS

Participants were required to be at least 18 years old, speak English or Spanish, self-identify as HA and reside in San Diego County. Recruitment of the randomized community sample was designed for an approximately even sample of Spanish- and English-speaking women.

Five-hundred three (\(N = 503\)) HA women were recruited from San Diego County in from October 1997 to February 2008. Half of the women (\(n = 251\)) reported a preference for materials and discussion in Spanish, and thus completed questionnaires in Spanish. Half (\(n = 252\)) reported English-language preference, and completed questionnaires in English. Participants’ ages ranged from 18-80 years with a mean age of 38.18 years (\(SD = 3.2\)).

MEASURES

All participants completed the DUREL either in English or in Spanish. The Spanish version was developed previously by Marcelo Calero. The five items and item responses for each version are listed in Table 1. In the English version, all five items are reverse coded. In the Spanish version, only Item Two is reverse coded. Adding the scores from Items Three, Four and Five (reverse coded for the English version) provides the score for the IR subscale. Total IR score can range from three to fifteen (Sherman et al., 2000). Summing the score for each item (reverse scored for English items and for Item Two in the Spanish version) generates a composite score of religiosity (Koenig et al., 1997), although, as noted above, more recently Koenig and colleagues have recommended against use of a total score (Koenig & Büssing, 2010). The total score can range from five to twenty-seven (Koenig et al., 1997; Sherman et al., 2000).

PROCEDURES

Recruitment and data collection followed the protocol outline below.
Table 1. Duke University Religion Index, English and Spanish Versions

<table>
<thead>
<tr>
<th>English</th>
<th>Spanish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How often do you attend church or other religious meetings?</td>
<td>1. ¿Con cuánta frecuencia asiste usted a la iglesia u otros encuentros religiosos?</td>
</tr>
<tr>
<td>1. More than once/wk</td>
<td>1. Nunca</td>
</tr>
<tr>
<td>2. Once a week</td>
<td>2. Una vez al año o menos</td>
</tr>
<tr>
<td>3. A few times a month</td>
<td>3. Unas pocas veces al año</td>
</tr>
<tr>
<td>4. A few times a year</td>
<td>4. Unas pocas veces al mes</td>
</tr>
<tr>
<td>5. Once a year or less</td>
<td>5. Una vez a la semana</td>
</tr>
<tr>
<td>6. Never</td>
<td>6. Más de una vez a la semana</td>
</tr>
<tr>
<td>2. How often do you spend time in private religious activities, such as prayer, meditation or Bible study?</td>
<td>2. ¿Con cuánta frecuencia dedica usted tiempo a actividades religiosas privadas, como rezar, meditar, o estudiar la Biblia?</td>
</tr>
<tr>
<td>1. More than once a day</td>
<td>1. Mas de una vez al día</td>
</tr>
<tr>
<td>2. Daily</td>
<td>2. Diariamente</td>
</tr>
<tr>
<td>3. Two or more times/week</td>
<td>3. Dos o más de dos veces al día</td>
</tr>
<tr>
<td>4. Once a week</td>
<td>4. Una vez a la semana</td>
</tr>
<tr>
<td>5. A few times a month</td>
<td>5. Unas pocas veces al mes</td>
</tr>
<tr>
<td>6. Rarely or never</td>
<td>6. Raras veces o nunca</td>
</tr>
<tr>
<td>3. In my life, I experience the presence of the Divine (i.e., God).</td>
<td>3. En mi vida, siento la presencia del Divino (i.e., Dios).</td>
</tr>
<tr>
<td>1. Definitely true of me</td>
<td>1. Definitivamente cierto para mí</td>
</tr>
<tr>
<td>2. Tends to be true</td>
<td>2. Tiende a ser cierto</td>
</tr>
<tr>
<td>3. Unsure</td>
<td>3. No estoy seguro</td>
</tr>
<tr>
<td>4. Tends not to be true</td>
<td>4. Tiende a no ser cierto</td>
</tr>
<tr>
<td>5. Definitely not true</td>
<td>5. Definitivamente no es cierto</td>
</tr>
<tr>
<td>4. My religious beliefs are what really lie behind my whole approach to life.</td>
<td>4. Mis creencias religiosas son lo que realmente está detrás de mi enfoque hacia la vida.</td>
</tr>
<tr>
<td>1. Definitely true of me</td>
<td>1. Definitivamente cierto para mí</td>
</tr>
<tr>
<td>2. Tends to be true</td>
<td>2. Tiende a ser cierto</td>
</tr>
<tr>
<td>3. Unsure</td>
<td>3. No estoy seguro</td>
</tr>
<tr>
<td>4. Tends not to be true</td>
<td>4. Tiende a no ser cierto</td>
</tr>
<tr>
<td>5. Definitely not true</td>
<td>5. Definitivamente no es cierto</td>
</tr>
<tr>
<td>5. I try hard to carry my religion over into all other dealings in life.</td>
<td>5. Trato de llevar mis fundamentos religiosos a todos los demás aspectos de mi vida.</td>
</tr>
<tr>
<td>1. Definitely true of me</td>
<td>1. Definitivamente cierto para mí</td>
</tr>
<tr>
<td>2. Tends to be true</td>
<td>2. Tiende a ser cierto</td>
</tr>
<tr>
<td>3. Unsure</td>
<td>3. No estoy seguro</td>
</tr>
<tr>
<td>4. Tends not to be true</td>
<td>4. Tiende a no ser cierto</td>
</tr>
<tr>
<td>5. Definitely not true</td>
<td>5. Definitivamente no es cierto</td>
</tr>
</tbody>
</table>
Recruitment

Recruiters used previously-tested community-based recruitment techniques. Women were recruited during meetings at community sites, by placing fliers in community-based sites (such as health clinics, churches and workplaces), and by word-of-mouth strategies. A key approach to reaching Mexican-American women and their families is to go through community channels, given that Mexican-Americans most often receive service-related information through informal sources, such as family, friends, and neighbors (Padilla & Villalobos, 2007). Staff from Vista Community Clinic’s Health Promotion Center recruited 63.5% of the participants (Vista Community Clinic is located in Vista, California); 36.5% were recruited by UCSD/Moore’s Cancer Center Community Outreach Department (UCSD/Moores Cancer Center is located in La Jolla, California). Recruiters recorded the method by which the participant was recruited, the number of eligible women recruited at each location, refusal rates, reasons for refusal and additional potential contacts.

Recruiters informed participants that their participation would involve completing questionnaires used frequently in health and research settings in order to learn if the questions worked well for HA women. Recruiters offered each participant the option of having the forms and surveys read aloud to her, in English or in Spanish, or of filling out the forms herself. Each participant came voluntarily to a location of her convenience (e.g., community clinic, home, workplace, church, community center) where, in a quiet setting, she completed written informed consent documents followed by a questionnaire packet, either individually or in small groups.

Data Collection

Participants completed the DUREL within a questionnaire packet, consisting of 14 scales. The DUREL was the sixth scale administered.

Seven female bilingual (English and Spanish) data collectors (four Mexican-American, two white American and one black American) collected the data from October 2007 through February 2008. All data collectors were IRB- and HIPAA-certified through University of California, San Diego (UCSD) online tutorials. The UCSD research team had trained the data collectors in survey data collection methods (e.g., practices, procedures,
protocol for screening for eligibility, confidentiality, informed consent procedures, reliable and thorough data collection, and reading of standardized scales to participants if needed).

The welcome, consent, data collection and subsequent question/discussion process took approximately one hour for participants who completed the questionnaires themselves. The process took about two hours for participants who elected to have the interviewer read the materials aloud. Each participant was offered light refreshments and was compensated with twenty dollars ($20.00) for her contribution and time.

The data from this collection are securely stored within UCSD Moore’s Cancer Center. The funding for the data collection came from the California Breast Cancer Research Program. UCSD Moore’s Cancer Center approved the use of this data for the present examination, and gave permission for Patricia Wansley Taylor (author/Master’s student defending this thesis) to use the data. The Institutional Review Board at San Diego State University approved the research protocol for the present study.
DATA ANALYSIS AND DETERMINATION OF MODEL FIT

Our approach to analyzing the data and determining whether a model fit the data is described below.

OBJECTIVE ONE: PROVIDING NORMATIVE DATA FOR THE DUREL

To provide normative data for the English and Spanish language versions of the DUREL in HA women, mean scores, standard deviations, and ranges were calculated and summarized separately for the HA-T group of women, and for HA-E and HA-S subgroups. SPSS version 17.0 was used for descriptive analyses.


We examined the reliability and factor structure of the DUREL as follows. One-factor and three-factor models were considered.

Investigating a Three-Factor Model (OR, NOR, IR) in EFA

The DUREL was developed on a model of three factors (OR, NOR, and IR) that theoretically comprise one overarching latent variable of “religiosity.” OR is derived from responses to Item One of the DUREL, NOR from responses to Item Two, and the IR from Items Three, Four and Five. A one-factor model (see Figure 1) has been demonstrated reliably to fit data from several demographic groups. We continued the investigation of the psychometric properties of the DUREL by testing and comparing a theoretical model postulating a simple single factor, “religiosity” (See Figure 2) to one postulating three correlated factors (OR, NOR, IR) underlying one larger factor (religiosity). We investigated
Figure 1. One-factor model of religiosity for DUREL scale.

Figure 2. Three-factor model of religiosity for DUREL scale.
each of the models in a sample of HA females \( (N = 503) \) with two subgroups. The HA-E subgroup chose to complete the measure in English using an English version of the DUREL \( (n = 252) \). The HA-S subgroup chose to complete the measure in Spanish using a Spanish version of the DUREL \( (n = 251) \). The DUREL was previously unexamined in groups with these demographic characteristics. We expected to see evidence for high internal consistency, and three distinct but related factors under one larger umbrella factor of religiosity. We used the analytic approach described as follows.

**Analytic Approach for Examination of Total Score and IR Score for Reliability and Factor Structure**

We examined reliability by calculating internal consistencies for the total score and the IR score using Cronbach’s alpha (see Cronbach, 1951). Next, we examined factor structure. We used exploratory factor analysis (EFA) to investigate the three-factor (OR, NOR, IR) model. EFA is a data reduction technique that assumes latent factors underlie observed variables. The purpose of EFA is to create fewer factors from the observed variables by factoring the relations among observed variables. We were interested in evidence as to whether the three factors contributed to an overarching latent variable of religious involvement or “religiosity.” To test for structural validity of the one-factor model, we used confirmatory factor analysis (CFA). CFA allows for a theory-driven study of the relationships between a set of observed variables and a set of continuous latent variables, and for comparison of competing models’ factor structures and model fit. We were unable to use CFA to test the three factor model due to limitations of a just-identified model, as mentioned earlier.

**Analytic Approach Using EFA for the Three-Factor Model**

We used (EFA) in SPSS 20.0 to investigate whether the number of factors that would be extracted from our observed variables would suggest a three-factor structure (OR, NOR, IR) underlying a one-factor structure. We first determined whether responses to DUREL Items Three, Four and Five from each of our three population groups suggested one factor (the latent variable of IR) using EFA. Next, we used the factor loadings generated from that
exploration to create a new variable for IR. We then ran EFA using our data for DUREL Item One, Item Two, and our newly created IR variable for each of our three groups.

**ANALYTIC APPROACH USING CFA FOR THE ONE-FACTOR MODEL AND THREE-FACTOR MODELS: ESTABLISHING MODEL FIT**

To evaluate the statistical significance of individual model parameters (e.g., factor loadings, factor variances, interfactor correlations), a statistical significance level of .05 was set. Tanaka (1993) suggests that chi-square tests may be insufficient in establishing model fit. Chi-square is influenced by sample size, so using a large sample size may lead to indication that no model fits (Bollen, 1986; Marsh, Balla, & McDonald, 1988). Consequently, we determined the best model fit for the data by the use of both statistical and descriptive indices. We determined overall model fit following the recommendations of Bentler (2007), using descriptive determinants in addition to the Chi-Square statistical determinant. The first descriptive determinant we used was the Root Mean Square Error of Approximation (RMSEA) (Steiger, 1990), with values less than .05 indicating a good model fit for the data, and values under .08 indicating acceptable fit. We also used the Comparative Fit Index (CFI) (Bentler, 1990), with values greater than .95 indicating that the model fits well, and values over .90 indicating that the model is plausible. Finally, we used the Standardized Root Mean Square Residual (SRMR) (Hu & Bentler, 1999), with values under .05 indicating the model is a good fit to the data, and values less than .08 indicating reasonable model fit. We considered the model a good fit to the data if it met two of the three descriptive criteria.

Mardia’s coefficient normalized estimate for the entire sample was 6.15. Mardia’s for the HA-E subgroup (7.13) and the HA-S subgroup (20.83) revealed multivariate nonnormality (Mardia, 1970; Satorra & Bentler, 2001). Data departing from multivariate normality may inflate the $\chi^2$ maximum likelihood ratio. In this case, the non-normality related to sample size and normality for the sample was acceptable without correction.

**OBJECTIVE FOUR: EVALUATION OF CROSS-LANGUAGE MEASURE INVARIANCE**

To substantiate that the measure retains its psychometric properties regardless of demographic difference, we further explored the DUREL’s psychometric properties by cross
sectional comparison. Multigroup CFA is the statistical method used to test measurement and structural invariance between groups. We employed one-factor multigroup CFAs, simultaneously estimating distinct models for the HA-E and the HA-S subgroups, while imposing equality constraints across the groups on relevant model parameters. In MPlus 6, we fit three multigroup models to the data for the HA-E and HA-S participants that matched the one and three factor structures indicated by previous research (Bentler, 2004). Following Vandenberg and Lance (2000), we next considered the fit of three multigroup CFA models (configural invariance, metric invariance, factor variance invariance) to data collected with the DUREL, using the same fit indicators described above.

The least restrictive model is the configural invariance model. This model tests whether the HA-E and HA-S subgroups have the same factor structure across groups and does not impose equality constraints. This model provides the value used as the starting point for comparison with a more restrictive model. If the pattern of factor loadings for both groups indicates that the constructs are substantially the same across groups, this model will be compared to a more restrictive model (Vandenberg & Lance, 2000).

The metric invariance model is more restrictive. By restricting each item’s factor loading to equivalence across groups, this model tests whether each item of the Spanish and English preference groups loads onto each factor equivalently. If this model fits well, then regardless of language, the association between each item and its respective factor(s) can be considered the same. A poor fit suggests true differences in response patterns between group, or measurement bias. The results from this model will compared to the configural invariance model results using a \( \chi^2 \) difference test to determine if the metric invariance model shows improved fit. If so, this will determine that the metric invariance model fits the data better than the configural invariance model and we will proceed with a further restrictive model.

The most restrictive model is the factor variance invariance model. This model examines whether the factors have equivalent variability between language groups. If we determine this model fits the data better than the previous one, using the \( \chi^2 \) difference test, this indicates that each language group has the same range of scores. If this model fits the data worse than the previous model, this suggests that the range of scores in one group is smaller than the other, due to genuine differences between the groups or to measurement bias.
OBJECTIVE FIVE: EXAMINING RELIGIOSITY’S LINKS WITH AGE

To investigate the associations between religiosity and age, SPSS versions 17.0 and 20.0 were used for correlational analyses. Previous research indicates positive correlations between age and scores of religiosity, OR, NOR and IR. We expected significant positive correlations for all scales with age for our sample and the two sub-groups. To examine these associations for each group, we conducted Pearson correlations of age and scale score, using Bonferroni corrections for alpha ($\alpha = .05/k$, i.e. $\alpha = .05/3 = .017$) to reduce the chance of a Type I error, i.e., rejecting the null hypothesis when it is true (see Curtin & Schulz, 1998). Data for age came from participants’ answers to Personal Health Survey, Item One, which asked participants to give their date of birth. The Personal Health Survey was administered to participants in the same questionnaire packet as the DUREL.
RESULTS

The following sets forth the results of the analyses described above.

OBJECTIVE ONE: PROVIDING NORMATIVE DATA FOR THE DUREL

Fifteen of 503 questionnaires were returned with missing responses to one or more items. For the DUREL, one missing response accounts for 20% of the data from that participant. Therefore, we opted to exclude participants who had one or more missing responses from the analyses. This affected only 2.98% of participants. Thus, the total number of participant responses was 488. A total of 245 English-language surveys and 243 Spanish-language surveys were analyzed.

The means and standard deviations for each item on the DUREL are reported in Table 2. Each group showed the full range of possible scores for each item, i.e. scores from one to six for Items One and Two, and scores of one to five for Items Three, Four, and Five. Participants’ total DUREL scores ranged from five to 27 (\(M = 19.64, SD = 4.74\)). Mean total scores were significantly different between groups, \(t (486) = -4.67, p < .001\). Scores for HA-S participants tended to show higher mean scores with less variability than scores for HA-E participants. See Table 3.

Ages ranged from eighteen to eighty years old, with a mean age of 38.10 years (SD = 13.21, \(N = 484\)). Over half (52.3%) of the women reported that they were married, 21.3% had never been married, 10.5% lived with a partner, and the remainder were divorced, widowed, or separated. Parents of one or more children were 77.9% (\(N = 485\)) of the sample. Fifty-nine percent of respondents reported employment. Highest level of education completed was reported as less than a high school education by 42.8%, completion of high school by 18.4%, and further education past high school (vocational/trade school, undergraduate or graduate education) by 38.5% of respondents. Religious affiliation of Christian and/or Catholic was reported by 94.7% of respondents (see limitations discussed on page 37). Percentage of respondents reporting birth in Mexico was 58.8%, whereas birth in the U.S. was reported by 34.0% of respondents.
### Table 2. Item Response Descriptives for Sample of Hispanic Women and English and Spanish Subgroups

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Item</th>
<th>Mean Total</th>
<th>SD Total</th>
<th>Mean English-preference</th>
<th>SD English-preference</th>
<th>Mean Spanish-preference</th>
<th>SD Spanish-preference</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>Item 1: How often do you attend church or other religious meetings?</td>
<td>3.90</td>
<td>1.44</td>
<td>3.80</td>
<td>1.49</td>
<td>4.00</td>
<td>1.40</td>
</tr>
<tr>
<td>NOR</td>
<td>Item 2: How often do you spend time in private religious activities, such as prayer, meditation or Bible study?</td>
<td>3.05</td>
<td>1.80</td>
<td>2.83</td>
<td>1.83</td>
<td>3.27</td>
<td>1.74</td>
</tr>
<tr>
<td>IR</td>
<td>Item 3: In my life, I experience the presence of the divine (i.e., God).</td>
<td>4.47</td>
<td>.92</td>
<td>4.29</td>
<td>.98</td>
<td>4.65</td>
<td>.82</td>
</tr>
<tr>
<td></td>
<td>Item 4: My religious beliefs are what really lie behind my whole approach to life.</td>
<td>4.22</td>
<td>1.03</td>
<td>4.00</td>
<td>.99</td>
<td>4.45</td>
<td>1.01</td>
</tr>
<tr>
<td></td>
<td>Item 5: I try hard to carry my religion over into all other dealings in life.</td>
<td>4.00</td>
<td>.92</td>
<td>3.75</td>
<td>1.24</td>
<td>4.25</td>
<td>1.08</td>
</tr>
</tbody>
</table>

OR = Organizational Religiosity  
NOR = Non-Organizational Religiosity  
IR = Internal Religiosity
Table 3. Subscale Descriptives for Sample of Hispanic Women and English and Spanish Subgroups

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Total Hispanic Sample ($N = 488$)</th>
<th>English-preference Subgroup ($n = 245$)</th>
<th>Spanish-preference Subgroup ($n = 243$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR</td>
<td>Mean: 3.90 SD: 1.44</td>
<td>Mean: 3.80 SD: 1.49</td>
<td>Mean: 4.00 SD: 1.40</td>
</tr>
<tr>
<td>NOR</td>
<td>Mean: 3.05 SD: 1.80</td>
<td>Mean: 2.83 SD: 1.83</td>
<td>Mean: 3.27 SD: 1.74</td>
</tr>
<tr>
<td>IR</td>
<td>Mean: 12.69 SD: 2.69</td>
<td>Mean: 12.04 SD: 2.78</td>
<td>Mean: 12.04 SD: 2.78</td>
</tr>
</tbody>
</table>

OR = Organizational Religiosity  
NOR = Non-Organizational Religiosity  
IR = Internal Religiosity

**Objective Two: Investigation of Reliability**

For the five-item measure of religiosity, Cronbach’s alphas were calculated for the entire sample ($N = 488$), and the HA-E ($n = 245$), and HA-S ($n = 243$) subgroups. DUREL Items One and Two have an item response range of one to six, whereas Items Three, Four and Five use a range of one to five. Due to the mixed ranges, we used z-scores to calculate Cronbach’s alphas. Results indicated that total scores for the HA-T group and from the HA-E subgroup had good internal consistency ($\alpha = .77$ and $\alpha = .82$, respectively). However, total scores gathered from the HA-S subgroup were not as internally consistent ($\alpha = .66$).

On the three-item subscale measure of IR, Cronbach’s alphas were also calculated from z-scores for the HA-T group ($N = 488$). Results indicated high internal consistency ($\alpha = .93$). Internal consistency for the three-item measure of IR was likewise satisfactory for both English and Spanish preference versions ($\alpha = .82$ and $\alpha = .77$, respectively).

**Objective Three: Investigation of Structural Validity**

We continued by examining the factor structure of the DUREL and its IR subscale.
Exploring Factor Structure of IR Subscale (Items Three, Four, and Five)

We explored whether EFA suggested that DUREL Items Three, Four, and Five group together as one factor (i.e., the factor of IR). Results indicated that for the sample and both subgroups, these three components did load onto one factor (IR). In the HA-T group, a one factor solution accounted for 73.11% of the solution’s explained variance, with factor loadings of .88 (Item Five), .87 (Item Four), and .81 (Item Three). In the HA-E group, one factor accounted for 87.32% of explained variance, with factor loadings of .95 (Item Four), .93 (Item Three), and .93 (Item Five). In the HA-S group, a one-factor solution accounted for 69.01% of explained variance, with factor loadings of .87 (Item Five), .86 (Item Four), and .76 (Item Three).

Exploring a Three-Factor Structure in DUREL (OR, NOR, IR)

Based on the results suggesting that Items Three, Four and Five group as one factor (IR), we proceeded to investigate whether our data generated support for a three-factor model for the DUREL in our female HA populations. For the sample and both subgroups, EFA derived a one-factor solution (i.e., “religiosity”) from the three variables entered (scores for Item One, Item Two, and factor loadings from the EFA on Items Three, Four and Five), suggesting a model of three-factors underlying an overarching latent variable of religiosity.

For the HA-T group, the one-factor solution derived from our three subscales explained 63.55% of the variance, with factor loadings for OR, NOR, and IR of .84, .78, and .77, respectively. For the HA-E group, the one-factor solution derived from the three subscales explained 70.36% of the variance, with factor loadings of .85 (OR), .84 (NOR), and .83 (IR). For the HA-S group, a one-factor solution composed of our three subscale factors accounted for 54.59% of explained variance, with factor loadings of .84 (OR), .69 (NOR), and .68 (IR). Across all groups, each of the three factors (subscales) contributed over 13% of explained variance toward the overall solution.

Confirming the One-Factor Model

We further examined the fit of the data to a one-factor model with the latent variable designated as “religiosity.” In this model, five observed variables, which are the responses to
the five items of the DUREL, contribute directly to the latent variable of religiosity (see Figure 1). We used confirmatory factor analysis (CFA) in MPlus 6 to examine the model’s fit. The factor loadings were estimated freely in this model, i.e. we did not constrain parameter estimates to equality across groups.

The data from the HA-T group did not fit a single factor model well statistically ($\chi^2 [5, N = 488] = 68.24, p < .001$). Descriptively, the model fit reasonably well, according to CFI (.92) and SRMR (.06), though RMSEA (.16) did not indicate a good fit. The results provide evidence for a single factor model. All standardized factor loadings were significant at the level of $p < .001$, with values ranging from .45 (Item Two) to .83 (Item Four). The factor loadings offer additional evidence for a one-factor model.

We examined the data’s fit to a simple single factor solution by conducting the same analysis for configural invariance on data from the HA-E and HA-S subgroups. For HA-E, the data did not fit the single-factor model well statistically ($\chi^2 [5, n = 245] = 34.95, p < .001$), but was a plausible fit descriptively, according to CFI (.94), and SRMR (.05), although not according to RMSEA (.16). Standardized factor loadings for the HA-E subgroup were generally large (values ranged from .63 (Item Two) to .85 (Item Four) and all were significant at the $p < .001$ level, offering further support for a single-factor model. However, the single-factor model was not plausible for the HA-S subgroup. The model was a poor fit to the data statistically ($\chi^2 [5, n = 243] = 40.46, p < .001$), and descriptively (CFI = .88, SRMR = .07, RMSEA = .17). Standardized factor loading values ranged from .21 (Item Two) to .81 (Item Four). The standardized factor loadings were statistically significant at $p < .001$, except for Item Two, which was significant at $p < .01$. Results did not establish configural invariance across groups. Therefore, we discontinued our evaluation with more restrictive models.

**Objective Four: Evaluation of Cross-Language Measure Invariance**

For HA-E, we found support for both a one-factor structure and a structure of three factors underlying a single latent variable. However, for HA-S, we only found support for the structure of three-factors contributing to a larger latent factor, which solution accounted for 15.77% less variance than it did for HA-E. Therefore, we were unable to confirm the same factor structure for HA-E and HA-S.
Objective Five: Examining Religiosity’s Links with Age

We predicted that increased age would relate positively to increased religiosity ($\alpha = .05/3 = .017$). Our data confirmed this prediction. For HA-T, age was significantly and positively associated with DUREL total score ($r = .23, p < .001, N = 484$). Age was also significantly and positively associated with DUREL total score in both HA-E ($r = .23, p < .001, n = 244$) and HA-S ($r = .18, p < .001, n = 240$).

We anticipated a significant positive correlation ($\alpha = .05/3 = .0166$) between age and OR and found it in two of our groups. For HA-T, there was a significant positive association between age and OR ($r = .12, p < .017, N = 484$). For HA-E, a stronger positive correlation was also found ($r = .20, p < .01, n = 244$). However, no correlation between age and OR was found for HA-S participants ($r = -.004, p > .05, n = 240$).

We found positive relationships between aging and NOR, as expected. For the HA-T sample, NOR significantly and positively associated with age ($r = .21, p < .001, N = 484$). Similar significant positive relationships were also found for HA-E ($r = .17, p < .01, n = 244$) and HA-S ($r = .22, p < .001, n = 240$).

IR significantly and positively correlated with age for HA-T ($r = .21, p < .001, N = 484$), and for HA-E ($r = .20, p < .01, n = 244$), as hypothesized, whereas HA-S had a positive, but nonsignificant relationship between age and IR ($r = .14, p < .17, n = 240$).
DISCUSSION

For HA women, the English and the Spanish versions of the DUREL do not
demonstrate an equivalent factor structure. The DUREL appears to have better reliability and
structural validity for more acculturated HA women (i.e. who chose to take the DUREL in
English, than for less acculturated HA women (i.e. those with a Spanish preference). We
found support for a model of three overlapping but distinct factors under a larger overarching
factor of religiosity (see Figure 2) for the total group and the two subgroups. However, we
were unable to confirm equivalence for the simpler one-factor model (see Figure 1) across
groups, which indicates genuine differences in response patterns between the HA women
who chose to take the measure in English, and those who chose to take the measure in
Spanish.

NORMATIVE DATA FOR THE DUREL

Participants’ total DUREL scores showed the full range of possible scores for each
item. Mean total scores were significantly different between groups. Subscale scores of
participants with a Spanish language preference tended to show higher mean scores with less
variance than participants with an English preference (see Table 3).

RELIABILITY

The DUREL appears to have good internal consistency when used with HA women
of mixed Spanish and English preference and with HA women with English preference,
which is consistent with previous studies (Koenig et al., 1997; Sherman et al., 2000; Sherman
et al., 2001). However, we found the Spanish version had less internal consistency.
Consistent with previous research (Koenig et al., 1997; Sherman et al., 2000; Sherman et al.,
2001), the IR subscale of the DUREL showed satisfactory internal consistency for all three
groups.
EXAMINATION OF THE FACTOR STRUCTURE AND OF CROSS-LANGUAGE MEASURE INVARIANCE

We further examined the data gathered from the DUREL to investigate its factor structure and to determine whether the data indicated cross-language measure invariance.

Factor Structure Using EFA

We used EFA to explore the factor structure of the three components of the IR subscale for our three sample groups. Results suggested a one-factor solution for the three components of the IR subscale.

We were unable to use confirmatory factor analysis (CFA) to test for a three-factor model (OR, NOR, IR), by specifying items to subscales, because the DUREL uses only one item for two of the three theorized factors, or by using total subscale score, because this would be a just-identified model. Turning to exploratory factor analysis (EFA), our results support a three-factor solution (OR, NOR, IR) underlying one overarching latent variable of “religiosity.” Our findings align with the design of the DUREL to measure the latent factor of religiosity by capturing the three prominent dimensions of religious involvement (organizational, non-organizational, and intrinsic dimensions of religiosity) that are central to any study of religiousness, and are often linked to health outcomes (Klemmack et al., 2007; Koenig & Büssing, 2010; Koenig et al., 1997; Storch, Roberti, Bravata, et al., 2004; Storch, Roberti, Heidgerken, et al., 2004).

We found high correlations among the three variables (OR, NOR, IR), suggesting an overlap in the factors and a common latent variable. This is consistent with the work of Sherman and colleagues (2000), who examined correlations among the subscales of the DUREL for discriminant validity and found significant overlap among the scales with moderate correlations. We found higher multicollinearity than Sherman and colleagues (2000), suggesting a strong overlap between DUREL subscales for HA women.

For meaningful interpretation of relationships with health outcomes, Koenig and Büssing (2010), and Sherman and colleagues (2000) recommend the use of subscale scores, and against the use of a total score. However, the high intercorrelations found in our data and in previous research (Storch, Roberti, Heidgerken, et al., 2004) suggest the DUREL captures one overall factor along with three interrelated sub-factors. A total score may be useful in investigating religiosity in situations unrelated to health outcomes, as well as for
investigating the relationship of the larger factor of religiosity (total score) with health outcomes along with the examination of relationships of different facets of religiosity (subscales) with health outcomes. Recognizing the dynamic, multi-dimensional character of religious and spiritual involvement, Sherman and colleagues (2001) warned that combining divergent components of religiosity into a single total score could obscure the different aspects of religiosity. However, this is different from not using a total scale score for any purpose.

**Factor Structure Using CFA**

Although our EFAs suggested a one-factor solution for all three samples, using CFA, for the Spanish preference group, we could not confirm a one-factor model (see Figure 1) using CFA. In contrast, data from the total HA group and for the English preference group did fit a one-factor model. Our results suggest an actual difference in response patterns between language groups, and that factor structure of the DUREL may be different across language or cultural groups.

**Cross-Language Measure Invariance**

We are thus unable to lend support for cross-language invariance between the Spanish and English versions of the DUREL. Differences in response alternatives, errors in translation, sampling bias, or Type II error may have influenced data from our Spanish preference sample. On the other hand, the response pattern diversity in our samples of HA women may represent an as yet unidentified factor or an interaction effect, such as cultural differences influencing ideation and behavior.

In sum, support for three overlapping but distinct factors within a larger factor of religiosity may best explain the results of our analyses of factor structure. We recommend further investigation to confirm or supplement our results.

**Religiosity and Age**

For our total group and English preference samples, age was positively related to total scores and each subscale scores of religiosity (OR, NOR and IR). For the Spanish preference sample, age was positively related to non-organizational religiosity, but did not relate significantly to organizational religiosity or intrinsic religiosity.
Our findings indicate that older HA women as an overall group, and HA women with an English preference, are more likely to have greater religiosity, greater frequency of attendance at religious services and meetings, greater frequency of private religious and spiritual practices, and more often use their religious commitments and beliefs as an orientation to life and to guide life decisions than do their younger counterparts. Results suggest that older HA women are more likely to attend religious services than younger HA women. However, for HA women with Spanish language preference, i.e. less acculturated HA women, age does not relate to the frequency of attendance at religious services or to religious commitment in daily life.

CONCLUSIONS

This study contributes to literature on the DUREL’s psychometric properties by exploring the reliability and structural construct validity among HA women in English and in Spanish. However, the implication of these findings for clinical and research contexts is that HA women with a Spanish language preference may not be responding to the measure in the same way as HA women with a preference for measures in English. After testing for configural invariance in each group, we discovered that the response pattern for HA women with a preference for forms in Spanish (less acculturated HA women) was not equivalent to that of English preference women (more acculturated HA women). These findings call for caution in using the DUREL with HA women in the U.S., especially for those with a Spanish preference (i.e. less acculturated HA women), and a call for further research.

LIMITATIONS OF THE PRESENT STUDY

This study has several limitations.

Demographic Limitations

This study has several limitations. All participants surveyed were located in San Diego County, and the majority (66.0%) were foreign-born, with women born in Mexico contributing a majority (58.8%) of the data. These characteristics limit generalizing our findings to HA women of other ethnicities and cultures.
Survey Limitations

The Spanish translation of the DUREL does not appear to be carefully and rigorously translated, thus creating excessive “noise” in the resulting analysis. One example is that response order, whether forward or reverse-coded, is not consistent between Spanish and English versions. As discussed by Schwartz (1999), respondents draw on response alternatives and preceding questions to make inferences about how they answer and understand a survey question. Therefore, differences in response order between versions are likely to increase variance in the resulting data.

Perhaps the most egregious translation error is response option three under Item Two in the Spanish version, which provides an incorrect translation of the DUREL’s English version response. The English version response is “Two or more times/week” and the supposedly equivalent Spanish version response is “Dos o mas de dos veces al dia,” i.e., two or more than two times per day. Moreover, this caused Spanish-preference participants to choose between “Mas de una vez al dia” (more than one time per day, which was the first response option) and “Dos o mas de dos veces al dia” (two or more than two times per day, the third response option).

Another limitation is that data on recruitment location were not collected. Given that recruitment sites included religious centers, participants contacted at these locations may have higher scores on the religiosity scales. Without data on recruitment, we cannot examine whether a disproportionate amount of participants were recruited from religious sites or whether data from participants recruited at religious sites showed greater religiosity than data collected at other community locations.

The percentage of our respondents who identified as Catholic and/or Christian was 94.7%, which may limit generalizability to HA women of other religious affiliations. It is unfortunate that an error in item response possibilities forced the Catholic respondents to choose between the responses of Catholic or Christian (i.e., in response to “What is your religious preference?” respondents chose between Christian, Catholic, Jewish et al.). Though 28.1% of participants chose “Christian” and 94.7% chose “Catholic”, we cannot interpret actual percentages due to this error in data collection.

According to PEW Research Center (2008), 58% of Hispanics identify as Catholic and 24% are Protestant (16% in the Evangelical tradition and 8% with other Protestant
denominations). For our purposes, we can consider that these numbers are likely higher, given that men are significantly more likely to report no religious affiliation than are women. Also relevant to our discussion are survey reports indicating that a large percentage of born in the U.S. who identified with Catholicism as a child, no longer identify as Catholic, whereas greater numbers of immigrants tend to identify as Catholic. In fact, 23% of U.S. Catholics are foreign-born and 82% of those are from Latin America and the Caribbean, which have very high Hispanic populations. Additionally, although a majority of HAs in the U.S. identify as Catholic, this percentage is decreasing (Newport, 2013). Younger HAs in the U.S. are less likely to be Catholic than those who are older. In contrast, the percentage of HA Protestants is consistent across age groups and has remained stable since 2008. This is important because, though HA Protestants are a smaller group, HA Protestants are much more likely to be very religious (60%) (responding that religion is an important part of their daily life and that they attend religious services every week or almost every week) than their Catholic counterparts (43%).

This is particularly important to our results not only because more U.S. Protestants (51%) are very religious, compared to 43% of U.S. Catholics, but because among HAs, this difference increases by seventeen percent (Newport, 2013). As a group, HA Protestants tend to be much more religious than the general population of U.S. Protestants, whereas HA Catholics tend to be no more religious than U.S. Catholics overall. This difference in religiosity has been found between HA Protestants and Catholics of all age groups.

**FUTURE STUDY**

Further research is needed to increase certainty of the psychometric properties of the Spanish version of the DUREL. Continued investigation of the DUREL’s reliability and factor structure with HA men and women representing particular ethnicities or countries of origin (e.g., Cubans, Puerto Ricans, Guatemalans, Spanish, etc.), in and outside of the U.S. is needed. If the measure is to be used in more than one language version within a population (e.g., English and Spanish for the HA population in the U.S. as in the present study), researchers investigating the factor structure and validity of DUREL translations should use caution when comparing scores between language versions. Future study is also needed to
determine whether and under what conditions using the DUREL total scale score is meaningful to any healthcare outcomes and to situations unrelated to health outcomes.

**RECOMMENDATIONS**

To increase the ease of analysis and interpretation, we recommend internal standardization of the DUREL scales and translated versions to facilitate analysis. Specifically, we recommend a uniform Likert-like scale for all five items, and conforming the reverse-scored items so that they match across language versions.
REFERENCES


