

Introduction

During the 1960s and 1970s, the U.S. Government examined various social indicators such as education and employment to better understand the quality of life (QOL) of its citizens (Bufford et al., 1991). Early QOL assessments focused on tangible experiences like physical well-being, standard of living, and health-related matters (Bufford et al., 1991; Paloutzian & Ellison, 2009; Scott et al., 1998). Examining physical and social indicators was thought to provide sufficient assessment of QOL; however, researchers soon realized that subjective experiences (e.g., happiness, spirituality) also contributed to this construct (Bufford et al., 1991). Within the counseling profession, spirituality has been an important construct not only in understanding one's QOL, but also as an integral component to the holistic approach to treatment.

Extensive counseling literature exists examining spirituality. Researchers have emphasized that spirituality and religion are increasingly important areas to research within the helping professions and counseling specifically (see Hall et al., 2004; Miller & Thoresen, 2003; Nelson et al., 2011; Prest et al., 1999; Sink & Devlin, 2011). The counseling profession has also identified spirituality as a core component of human development (Cashwell et al., 2010; Myers & Williard, 2003) and recognized its impact on assessment, intervention, and cultural and ethical issues (Sink & Devlin, 2011). Research has stressed the importance of spiritual competencies in training counselors to work with clients presenting with religious and/or spiritual issues (Cashwell & Young, 2004; Young et al., 2007). Researchers have also examined how counseling students feel about discussing and exploring spiritual issues with clients (Souza, 2002), the spiritual beliefs of current mental health professionals and how they use spirituality as an intervention (Hall et al., 2004; Morrison et al., 2009; Morrison & Borgen, 2010; Young et al., 2007), and the connection between counseling student's spiritual well-being and sense of calling to the profession (Hall et

al., 2014). Further, it has been suggested that counselors who value spirituality in their personal lives and view faith as desirable (Morrison & Borgen, 2010) are more likely to believe in a personal God than other helping professionals (Sheridan et al., 1992).

Although much attention in the counseling literature has appropriately been focused on the necessity and methodology of counselor training in spirituality, no known research has examined the construct validity of spiritual well-being using the *Spiritual Well-Being Scale* (SWBS; Paloutzian & Ellison, 1982). Much of the extant research on construct validity of spiritual well-being was conducted using the SWBS with non-counseling populations. Further exploration of the spiritual well-being construct is needed to better understand its validity and utility, especially since well-being is a fundamental construct shaping the counseling profession.

Spiritual Well-Being Scale

A number of spirituality measures exist throughout the literature (see Koenig, 2008; MacDonald & Friedman, 2002). A commonly used QOL assessment developed to measure one's subjective state of spiritual well-being is the SWBS. Koenig (2008) reported that the SWBS "is perhaps the most widely used spiritual well-being scale used today to assess spirituality" (p. 352) with over 182 articles. The SWBS measures one's subjective understanding of well-being and perceived spiritual QOL, as understood from a religious and existential perspective; however, the SWBS is not based on a specific religion or orientation. The SWBS was designed as a general measure of perceived well-being and has application for both individual and group assessment (Paloutzian & Ellison, 2009). In developing the SWBS, Paloutzian and Ellison believed that little attention was paid to the spiritual and existential aspects of people's lives (Scott et al., 1998). Since its inception, the SWBS has been used to conduct research and evaluation in various domains including nursing, healthcare, clinical work, and congregational assessment (Paloutzian & Ellison, 2009).

Influencing Paloutzian and Ellison's (1982) development of the SWBS was Moberg and Brusek's (1978) conceptualization that spiritual well-being has both vertical (i.e., connection with God) and horizontal (i.e., well-being is rooted in life purpose and meaning) dimensions. Paloutzian and Ellison (1982) developed the SWBS using a sample of 206 White American college students from four Western U.S. universities (three of which were religiously oriented). The SWBS is a 20 item self-assessment composed of two subscales, *Religious Well-Being* (RWB) referred to as the vertical dimension, and *Existential Well-Being* (EWB) measuring the horizontal dimension. An overall measure of *Spiritual Well-Being* (SWB) is assessed by the summated score of combined subscales. The vertical dimension items contain the word "God" and indicate the spiritual relationship that one has with their god. For example, "I believe that God loves and cares about me." The items on the horizontal dimension contain statements about life satisfaction and one's adjustment to self and their surroundings. For example, "I feel very fulfilled and satisfied with life."

The EWB and RWB consists of the even and odd number items, respectively (Paloutzian & Ellison, 2009). Items are measured using a six-point Likert scale. Positively worded items are numbered 3, 4, 7, 8, 10, 11, 14, 15, 17, 19, and 20, with possible scores ranging from 1 "Strongly Disagree" to 6 "Strongly Agree." Negatively worded items are numbered 1, 2, 5, 6, 9, 12, 13, 16, and 18, with possible scores ranging from 1 "Strongly Agree" to 6 "Strongly Disagree" and are reversed scored. Total scores for each of the subscales can range from 10-60. Total scores for the SWB can range from 20 to 120. Low scores (21-40) indicate a low sense of spiritual well-being whereas higher scores (100-120) indicate a high sense of spiritual well-being (Paloutzian & Ellison, 2009).

Reliability and Validity

Various studies demonstrated the SWBS is a reliable and valid measure. Paloutzian and Ellison (1982) examined test-retest reliability coefficients by administering the SWBS with a one-week interval. They identified reliability coefficients of .86 for the EWB, .96 for the RWB, and .93 for the SWB. Bufford et al. (1991) reported test-retest reliability data for three studies with retesting occurring from 4-10 weeks and reliability coefficients ranging from .73 to .99. Additionally, several studies using the SWBS found evidence of internal consistency with coefficients ranging from .82 to .97 for the RWB, .78 to .91 for the EWB, and .76 to .94 for the SWB (Bufford et al., 1991; Fernander et al., 2004; Genia, 2001; Hall et al., 2014; Paloutzian & Ellison, 1982). Bufford et al. (1991) provided support for construct validity with SWB, RWB, and EWB items correlating positively with sense of purpose, emotional adjustment, physical health, and positive self-concept and negatively with poor health, lack of purpose in life, and emotional maladjustment. Bufford et al. (1991) reported that items clustered appropriately into the designated EWB and RWB subscales.

Fee and Ingram (2004) found construct validity was supported with RWB, EWB and SWB being positively correlated with life purpose, intrinsic religiosity, and self-esteem measures, and negatively correlated with measures of loneliness. As reported by Utsey et al., (2005), the SWB was also correlated with life satisfaction and emotional well-being, ethnicity, depression and self-esteem. Concurrent validity was supported by the positive correlation between the SWBS and Crumbaugh and Maholick's *Purpose in Life Test* and several other measures (Fee & Ingram, 2004).

Psychometric Issues with the SWBS

Despite its frequent usage, researchers have raised concerns about the psychometric validity of the SWBS. Several researchers questioned whether the SWBS accurately measures

perceptions of spiritual well-being due to problems with ceiling effects, especially with evangelical populations (Bufford et al., 1991; Genia, 2001; Ledbetter, Smith, Fischer, et al., 1991; Scott et al., 1998). Examination of evangelical samples revealed data were substantially negatively skewed, highlighting concerns regarding the validity of the correlational analysis due to nonnormality distributions and violating assumption of equal variances (Bufford et al., 1991; Ledbetter, Smith, Volser-Hunter, et al., 1991; Scott et al., 1998). Further, when data are highly skewed, greater variability in factor structure may result (Bufford et al., 1991). Genia (2001) examined ceiling effects in a spiritually heterogeneous sample of 211 college students. Results suggested that ceiling effects were associated with religious affiliation, specifically with Christian participants whose scores yielded substantially negative skewness on the SWBS. Evidence of ceiling effects were also found for RWB and EWB across various religious groups (Genia, 2001).

The SWBS consists of positively and negatively worded items that can introduce potential issues due to response bias (i.e., method effects). The intent in using valence-based items is to minimize acquiescence, affirmation, or agreement bias, which can occur when participants tend to agree with the stated item regardless of the content (DeVellis, 2017). The assumption for positively and negatively worded items is that they are measuring the same content and should result in participants giving similar responses (Marsh, 1996). However, self-report measures that contain positively and negatively worded items may yield different factor structures, specifically structures that align with positively and negatively worded items (Carmines & Zeller, 1979; Tomás & Oliver, 1999). If valence-based items result in responses that differ from what was intended, then results may be confounded by a method effect due to biased data, thus calling into question the validity of results (DiStefano & Motl, 2006). A method effect occurs when the variance resulting from a

measurement procedure adds variance beyond what is attributed or expected from the variable or attribute itself (Maul, 2013).

Numerous studies have investigated construct validity of the SWBS. As mentioned, Paloutzian and Ellison (1982), using a sample of 206 students, reported the initial exploratory factor analysis (EFA) yielded a three-factor structure with factors labeled as Religious Well-Being (factor 1), Life Satisfaction (factor 2), and Life Direction (factor 3); however, Ellison (1983) indicated only factors 1 and 2 yielded eigenvalues greater than 1. Factor 3 failed to meet this criterion yet the items were retained. Paloutzian and Ellison (as cited by Scott et al.) reasoned that theoretical consistency factored into the decision to retain these items and integrate them into factor 2, resulting in the two-factor model. Scott et al. (1998) indicated the decision to retain items presents potential problems with construct and content validity and suggests that the items are measuring multiple constructs.

Ellison (1983) used an orthogonal rotation for the initial EFA, despite an association between RWB and EWB subscales. Several researchers suggested an oblique rotation (e.g., oblimin) may be more appropriate since factors appeared to be correlated (Ledbetter, Smith, Fischer, et al., 1991; Scott et al. 1998). Costello and Osborne (2005) mentioned that in the social sciences, factors are rarely unrelated, and one can generally expect an association exists between items so using orthogonal rotation when items are correlated may limit the information generated from the factors.

You and Yoo (2016) sampled 470 Korean adults from various settings to conduct an EFA and confirmatory factor analysis (CFA) to test the validity of the SWBS. Examining the nine negatively worded items as part of the EFA, they identified several cross-loaded items raising concern about method effect. One model included in the CFA was a two-factor model with method

effects to control for the problem cross-loading of negatively worded items. You and Yoo confirmed the two-factor model with method effects yielded a better fit than the one-factor or original two-factor model.

Scott et al. (1998) investigated the clinical value of the SWBS using archival data of 202 inpatient psychiatric clients. They conducted an EFA which posited a three-factor solution and found no significant difference between orthogonal and oblique factor solutions. One concern with this study was cross-loaded items, yet no criteria was stated to justify retaining them. For example, SWB12 and SWB18 cross-loaded highly on factors 2 and 3. Since these factor loadings were above the significance threshold of .40 for sample size of 200 (Hair et al., 1998), understanding their rationale and criteria for retaining items would be helpful.

Genia (2001) found both varimax and oblimin rotations supported the two-factor structure of the SWBS. Ledbetter, Smith, Fischer, et al. (1991) used archival data to conduct a CFA from two independent samples. Sample 1 included 285 subjects and sample 2 consisted of 319 subjects. Goodness-of-fit indicators revealed a poor fit for both the one- and two-factor models for each sample. Even though the two-factor model yielded a better model fit, results did not support a general single factor construct as theorized by Gorsuch (1984) or Paloutzian and Ellison's (1982) two-factor model suggesting that the SWBS may be factorially complex. Based on the correlations Ellison (1983) reported between the EWB and RWB, Ledbetter, Smith, Vosler-Hunter, et al. (1991) believed the SWBS might be better conceptualized as a one-factor construct.

Miller et al. (1998) conducted two separate EFAs, one with 119 White and a second with 97 African American undergraduate and graduate students. The purpose of this study was to examine whether perceptions of well-being were influenced by ethnicity. Their results postulated a three-factor structure for Whites and the five-factor structure for African Americans. Several

concerns were evident with this study including the: (a) small sample size for both groups, (b) use of orthogonal rotation, and (c) lack of clarity regarding criteria for retaining items (Utsey et al., 2005).

Utsey et al. (2005) conducted a confirmatory test of the SWBS with a community sample of 291 African Americans to determine the validity of the model fit established by Miller et al. The researchers made comparisons across identified SWBS factor models, including those using orthogonal and oblique rotation procedures and hierarchical models. Even though results of the CFA indicated that the five-factor model had reasonable fit, the authors concluded that none of the proposed factor solutions had acceptable model fit.

Fernander et al. (2004) conducted an EFA using a modified version of the SWBS with incarcerated male drug users. Data were collected from 631 Black and White incarcerated male drug users from four Kentucky correctional facilities. Data analysis supported a two-factor model with the RWB and EWB subscales consisting of 11 and 9 items, respectively. Item EWB10 was not included in either of the two-factor solutions. The authors used an orthogonal rotation method and found that White inmates scored significantly higher on both subscales than Black inmates.

As demonstrated, factorial validity for the two-factor model has not been fully supported. Ledbetter, Smith, Fischer, et al. (1991) emphasized the need for further examination of the factorial construct validity as they believe the SWBS is a factorially complex construct that is measuring multiple behaviors inconsistent with unidimensional subscales outlined by Paloutzian and Ellison. Additional research is also needed to address method effects due to negatively worded items (Ledbetter, Smith, Vosler-Hunter, et al., 1991; You & Yoo, 2016). The purpose of this study is to examine the psychometric qualities of the SWBS. Specifically, this study attempts to answer the following research questions: 1) Does the original two-factor model demonstrate better model fit

compared to hypothesized one, three, and five factor models? 2) Is there evidence of a method effect on negatively worded items?

Method

Participants

This study examined data from 415 master's level counseling students attending Council for Accreditation of Counseling and Related Educational Programs (CACREP) programs. Nearly 9 in 10 participants were female ($n = 367$). The average age of participants was 31 years ($SD = 9.98$) with age ranging from 21 to 65 years. Regarding program of study, 210 (51%) participants were enrolled in clinical mental health counseling, 100 (24%) were enrolled in school counseling, 89 (21%) were enrolled in community counseling, and 15 (4%) identified other counseling programs (i.e., rehabilitation, pastoral, and college counseling). For this study, rehabilitation, pastoral, and college counseling were collapsed into clinical mental health counseling due to the small number of participants in these three programs and similarities in populations served and issues addressed. The average number of credit hours completed across all programs was 28.50 ($SD = 20.43$) with credit hours completed ranging from 0 to 68. Mean score for SWB was 90.18 ($SD = 17.80$), EWB was 48.85 ($SD = 7.18$), and RWB was 41.33 ($SD = 14.75$). Internal consistency was .91 for the SWB, .97 for the RWB, and .86 for the EWB.

Procedure

This study adheres to American Counseling Association Code of Ethics and institutional review board approval was granted for this study. CACREP liaisons at all CACREP accredited master's programs in the United States ($N = 263$) were identified using a directory obtained from the CACREP website. E-mails were sent to each liaison containing a web link to an online survey that could be forwarded by the liaison to students currently enrolled in their program. Data were

collected using Survey Monkey. The survey also included a brief demographic questionnaire, including age, gender, program credits completed, and CACREP program of study.

Results

Pre-Analysis Data Screening

Sample Size

Prior to answering the research questions, pre-analysis data screening was undertaken. Before testing assumptions, an examination of missing data was conducted. The original study yielded 476 responses. When analyzing the pattern of missing data, 61 cases were determined to have a significant number of random, incomplete responses across the study variables. Due to the amount of missing data for these cases and the large sample size, imputation methods were not viewed as an appropriate solution. Consistent with Hair et al.'s (1998) recommendation, cases with excessive missing data across the study variables were deleted. After these cases were deleted a frequency report was run, and the analysis indicated no remaining missing data for this instrument. Regarding adequacy of sample size for CFA, a great deal of disparity exists within the literature (Wolf et al., 2013). Some researchers indicated the absolute number of cases rule is sufficient, while others indicated that sample size should be a ratio of subjects-to-variables. Comrey and Lee (1992) reported that total sample size of 300 is good and 500 is very good. Hair et al. (1998) mentioned that researchers should strive for the highest subject-to-variable ratio (e.g., 20:1), while minimizing overfitting the data. For this study, the sample size of 415 appears sufficient in terms of both absolute sample size and subject-to-variable ratio.

Model-fit Criteria. Determining model fit in CFA requires examination of several goodness-of-fit indices. One criterion is the examination of the model fit summary (CMIN/DF), the Likelihood Ratio Test statistic expressed as a chi-square (χ^2) statistic. Large sample sizes often

result in chi-square statistic yielding a statistically significant difference between the sample data and the hypothesized model. Since CFA is a large sample statistic, chi-square statistic is often a poor measure of model fit. Marsh and Hocevar (1985) recommended the Likelihood Ratio Test statistic be within the 2-5 range to indicate reasonable fit.

Given the limitations with chi-square to measure model fit, other statistics and indices were used to evaluate model fit. The Root Mean Square Error of Approximation (RMSEA) measures how well the hypothesized model would fit the population covariance matrix. RMSEA values $<.05$ are indicative of good fit and values $<.08$ indicate reasonable errors of approximation in the population (Byrne, 2016). Reporting the RMSEA 90% confidence interval assists in interpreting model fit, as well as closeness of fit measured by the probability value, PCLOSE, where values $>.05$ indicate a good fit with the population (Byrne, 2016). The Standardized Root Mean Square Residual (SRMR) represents the average residual between the hypothesized model and sample data, and the smaller the SRMR value the better with 0 indicating a perfect fit, $<.05$ reflecting a well-fitting model (Byrne, 2016), and $<.08$ is recognized as good fit (Hu & Bentler, 1999). Comparative Fit Index (CFI) accounts for sample size and provides a measure of complete covariation of the data, with values ranging from 0 to 1.00. Hu and Bentler (1999) recommend that a cutoff value of .95 is an indicator of better fit. Tucker-Lewis Index (TLI), which yields values between 0 to 1.00, is indicative of good fit when values are also close to .95 (Hu & Bentler, 1999).

Assumptions. Along with checking adequacy of sample size, assessment of multivariate outliers and assumption of nonnormality were examined. Regarding multivariate outliers, Mahalanobis distance (d^2) was reviewed for each case. No cases were viewed as distinctly different from others based on d^2 values (Byrne, 2016). The assumption of normality was also examined. Byrne (2016) indicated data that are multivariate kurtotic is problematic in CFA since kurtosis

affects tests of variances and covariances. Kurtosis values for the 20 SWBS items range from .385 to 2.59 for the positive values and -.124 to -1.13 for the negative values yielding an overall univariate mean kurtosis value of 1.40. West et al. (1995) mentioned that kurtosis values equal to or greater than 7 indicate departure from normality. For this sample, substantial kurtosis does not appear to be problematic. To determine if multivariate normality was of concern, Bentler (as cited in Byrne, 2016) suggested that a critical ratio (C.R.) value > 5.00 indicates multivariate nonnormality. In this sample, the C.R. value is 67.36 indicating strong likelihood of multivariate nonnormality. Attempts to address multivariate nonnormality using asymptotic distribution-free (ADF) estimation and bootstrapping revealed that neither procedure improved multivariate normality.

Data Analysis

Given the concerns with construct validity of the SWBS, we examined whether the confirmatory test supports Paloutzian and Ellison's (1982) two-factor model. We also examined if the two-factor model demonstrated better model fit compared to the hypothesized one-, three-, and five-factor models identified in the literature.

Table 1 highlights goodness-of-fit indices for the various models. Results indicate that none of the hypothesized models demonstrated good overall model fit. The one- and five-factor models reflected poor fit across all indices. The Scott et al. three-factor model fared better relative to the one- and five-factor models on all measures yet still resulted in poor model fit. The Paloutzian and Ellison two-factor (CMIN/DF = 3.59, RMSEA = .079, CFI = .93, TLI = .92) and Miller et al. three-factor (CMIN/DF = 3.39, RMSEA = .076, CFI = .94, TLI = .93) models had reasonable fit relative to the one- and five-factor models. As expected, given the large sample size, all chi-square values

were statistically significant. The SRMR values for the hypothesized two-factor and Miller et al. three-factor models reflected good fit relative to the one- and five-factor models.

Table 1

Goodness-of-fit Indices for SWBS Models

Model	χ^2	p	df	χ^2/df	RMSEA	RMSEA Lo90	RMSEA Hi90	RMSEA PCLOSE	CFI	TLI	SRMR
One-Factor	1768.70	.0001	170	10.40	.151	.145	.158	.0001	.74	.71	.176
One-Factor (Respecified)	843.24	.0001	158	5.34	.103	.096	.110	.0001	.89	.87	.144
Two-Factor (Paloutzian & Ellison)	606.15	.0001	169	3.59	.079	.073	.086	.0001	.93	.92	.078
Three-Factor (Miller et al.)	565.27	.0001	167	3.39	.076	.069	.083	.0001	.94	.93	.077
Three-Factor (Scott et al.)	573.46	.0001	101	5.68	.107	.098	.115	.0001	.91	.89	.097
Three-Factor (Scott et al. Respecified)	352.20	.0001	98	3.60	.080	.071	.089	.0001	.95	.94	.077
Five-Factor (Miller et al.)	1136.60	.0001	160	7.10	.122	.115	.129	.0001	.84	.81	.148
Five-Factor (Miller et al. Respecified)	820.10	.0001	157	5.22	.101	.095	.108	.0001	.89	.87	.127

χ^2 = chi-square, p = probability, df = degrees of freedom, χ^2/df = relative chi-square ratio, RMSEA = Root Mean Square Error of Approximation, RMSEA Lo90 = Lower limit of a 90% CI, RMSEA Hi90 = Upper limit of a 90% CI, RMSEA PCLOSE = p value for testing the H_0 population RMSEA, CFI = Comparative Fit Index, TLI = Tucker-Lewis Coefficient, SRMR = Standardized Root Mean Residual

As a result of poor model fit exhibited by the one-factor model, Scott et al. three-factor, and Miller et al. five-factor models to the sample data and at least two misspecified parameters within each model, post hoc model fitting seemed to be a reasonable and expected next step (Byrne, 2016). Modification indices (MI) and standardized residuals were examined for these models. Byrne (2016) indicated parameters with MI above 10 have significant influence on the overall model. Regarding residuals, several covariates had standardized residuals greater than 2.58 indicating possible concerns with model misfit (Byrne, 2016). As shown in Table 1, respecification of the one-factor model yielded some improvement in goodness-of-fit indices, however the resulting modifications did not improve the overall model fit.

Examining the Scott et al. three-factor model, all the parameters had MI above 10 with several covariates substantially different from the rest. EWB_5 “I feel a sense of well-being about the direction my life is head in” and EWB_6 “I don’t enjoy much about life” had the highest MI value. The respecified model yielded a large $\Delta\chi^2 = 106.43$ (Model 1 $\chi^2 = 573.46$ – Model 2 $\chi^2 = 467.03$) and an unstandardized error covariance estimate of .362, which is highly significant (C.R. = 8.723). Improvements in RMSEA (.095 vs. .107), CFI (.93 vs. .91), TLI (.91 vs. .89) were noted in model fit. Reviewing the MI, the next highest MI value was covariate EWB_1 “I don’t know who I am, where I came from, or where I am going” and EWB_2 “I feel that life is a positive experience” which was also substantially higher than the other items. The respecified model yielded continued improvement in model fit with $\Delta\chi^2 = 58.33$ (Model 2 $\chi^2 = 467.03$ – Model 3 $\chi^2 = 408.70$) and an unstandardized error covariance estimate of .552, which is highly significant (C.R. = 7.274). Slight improvements in RMSEA (.087 vs. .095), CFI (.94 vs. .93), TLI (.93 vs. .91) were noted. Item EWB_5 was specified as loading on the Factor 2 but the MI indicated that it should also load on Factor 3. The intent of this item, while not focused on aspects related to God,

has similarities with the other items in this construct regarding strength and meaningful direction of life and seemed appropriate to also load on Factor 3. Goodness-of-fit indices show improvement in model fit with the $\Delta\chi^2 = 56.50$ (Model 3 $\chi^2 = 408.70$ – Model 4 $\chi^2 = 352.20$), RMSEA (.080 vs. .087), CFI (.95 vs. .94), TLI (.94 vs. .93). The loading of item EWB_5 on both factors was both statistically significant and had similar critical ratios (6.441 vs. 4.224). Further, the estimated value for target loading (.457) is higher than it is for the cross-loading (.200) indicating moderate-to-low strength. The respecified Scott et al. three-factor model demonstrated reasonable model fit (and better model fit compared to the hypothesized model); however, respecification shifts the focus of the model to an exploratory mode raising concerns about its overall fit (Byrne, 2016; Harrington, 2009)

Post hoc model fitting was also conducted with the Miller et al. five-factor model. Model respecification was conducted despite concerns that the original model EFA had substantial flaws related to its methodology and data analysis (Utsey et al., 2005). As was the case with the one-factor model, respecification of the Miller et al. five-factor model showed some improvement in goodness-of-fit, however the modifications did not improve overall model fit.

The one and five-factor models and their respective respecified models demonstrated poor fit for this sample. The two and three-factor models presented better overall model fit compared to the hypothesized and respecified one- and five-factor models. The posited two-factor and Miller et al. three-factor models demonstrated better overall fit relative to Scott et al.'s three-factor model. The Scott et al. respecified three-factor model showed improvement across all indices compared to the hypothesized model. The Paloutzian and Ellison two-factor showed comparable results to the Miller et al. three-factor and Scott et al. respecified three-factor models, the latter two models

yielded better model fit and were generally comparable across all indices. As noted, caution should be exercised when interpreting respecification model fit.

We also examined the effect of the main construct (SWB) on the first-order factors. Our intent was to estimate the effect of the SWB on the first-order constructs identified to determine how well the theorized construct loaded into the underlying first-order constructs. In reviewing the SWBS factor validity studies identified in the literature, analysis of second-order CFAs for those studies indicating acceptable model fit was not evident. Second-order CFA is used to “confirm that the theorized construct in a study loads into certain number of underlying sub-constructs or components” (Awang, 2012, p.163). Second-order CFA is used when: a) items are theoretically related and b) first-order factors are based on a higher order factor structure (Schumacker & Lomax, 2010). We analyzed a second-order effect based on items being theoretically related to SWB. Of the three models identified, only the Miller et al. three-factor model was appropriate for analysis. The Paloutzian and Ellison two-factor model was not appropriate since a minimum of three first-order factors are needed to conduct higher order factor analysis (Kline, 2011). The Scott et al. three-factor respecified model also was also not appropriate since respecification shifts the model from validation to exploratory mode rejecting the hypothesized model (Byrne, 2016; Harrington, 2009). Model fit results for the second-order construct SWB yielded similar indices for model fit as the first-order model ($\chi^2 = 565.27$, $p < .000$, $\chi^2/df = 3.39$, RMSEA = .076, CFI = .94, TLI = .93, SRMR = .077). Thus, the second-order CFA is similar at reproducing the variables compared to the first-order model. Factor loadings for the SWB for the three factors in the Miller et al. second-order model were .20 for factor 1, 1.02 for factor 2, and .84 for factor 3. Standardized parameter loadings ranged from .77 to .93 for factor 1, .53 to .70 for factor 2 and .51 to .83 for factor 3.

We also investigated whether there was evidence of method effect on negatively worded items as reported by You and Yoo (2016). Specifically, did negatively worded items highly cross-load on Factors 1 or 2 during the EFA. Using a similar process outlined by You and Yoo, factor loadings from the hypothesized model were explored. No substantial cross-loadings were found on negatively worded items. Further, examination of the CFA models did not reveal a pattern of negatively worded items loading on any one factor. Response bias does not appear to be problematic with this sample.

Discussion

Spirituality has become an important QOL indicator in the counseling profession with researchers indicating greater emphasis is needed to understand this construct (Hall et al., 2004; Miller & Thoresen, 2003; Nelson et al., 2011; Prest et al., 1999; Sink & Devlin, 2011), especially with respect to assessment and intervention (Sink & Devlin, 2011). Despite the popularity of the SWBS, concerns about its psychometric validity persist. Several studies investigated the SWBS to establish its construct validity, yet little agreement exists regarding the best model fit, potentially creating difficulties in measuring or assessing spiritual well-being in counseling research. The purpose of the study was to further examine the construct validity of the SWBS.

Results for the first research question were inconsistent with findings throughout the literature. Overall, factorial validity for the hypothesized models did not demonstrate good fit. For this study, results were consistent with poor model fit of the one-factor model found by Utsey et al. (2005) and Ledbetter, Smith, Fischer et al.'s (1991) examination of archival data samples. Paloutzian and Ellison's posited two-factor model and Miller et al.'s three-factor model demonstrated better overall fit compared to the other hypothesized models. The two-factor model demonstrated adequate fit and supports the factor structure posited by Paloutzian and Ellison's EFA.

The RMSEA value for the current study's two-factor model was consistent with Utsey et al.'s findings for the oblimin and hierarchical two-factor models and showed better overall fit compared to Ledbetter, Smith, Fischer, et al.'s two-factor models.

The hypothesized Miller et al. three-factor model demonstrated comparable model fit to the two-factor model. As noted by Utsey et al. (2005), methodological and data analysis issues regarding the EFA conducted by Miller et al. raised concern about the veracity of the three-factor model, yet the goodness-of-fit indices appeared to support their three-factor model. Findings were not consistent with Scott et al.'s proposed three-factor model as evidenced by poor model fit. Scott et al. (1998) reported several highly cross-loaded items which raised questions about factorial validity. A post hoc analysis was conducted to determine if model fit could be improved. The Scott et al. respecified model showed improvement across all goodness-of-fit indices and yielded comparable model fit. However, as Byrne (2016) noted, post hoc analysis in structural equation modeling shifts the analyses to exploratory mode indicating the hypothesized model has been rejected and introduces the possibility of creating an overfitted model. Further analysis of this model is warranted to examine its validity.

Evidence of method effects due to negatively worded items did not support the findings from You and Yoo (2016). Even though some of the negatively worded items grouped together, there did not appear to be concerns with response bias for this sample. Previous research has suggested negative evaluation by others and lower levels of self-consciousness seem to have some influence on method effects with negatively worded items (DiStefano & Motl, 2006). Further, method effects with negatively worded items may be influenced by differing expectations males and females have about negative evaluation by others (DiStefano & Motl, 2009). With the sample of nearly all female participants, gender expectations may have contributed to lack of method

effect. Another possible reason method effects of negatively worded items were not evident may be a function of what the SWBS intended to measure. Further, the study did not focus on the influence of personality traits, sex differences as it relates to career selection, or cultural considerations which may explain the lack of response bias.

Implications for the Counseling Profession

Spiritual well-being is a core component underlying the holistic approach of counseling. Assessments that can accurately measure spiritual well-being enhance the theoretical understanding and practical application of this construct. Study results appear to support early findings that the SWBS may be factorial complex (Ledbetter, Smith, Fischer et al., 1991) possibly creating challenges for counselors attempting to measure spiritual wellbeing and its underlying constructs. One implication for counseling professionals who choose to use the SWBS is that more than two unique constructs may reflect spiritual well-being, consistent with several factorial validity studies. Further, since various iterations of model fit, drawn from the SWBS literature, were examined, and that no model demonstrated good fit for the SWBS may indicate some difficulty in interpreting spiritual well-being due to its multidimensionality.

Even though this study demonstrated adequate fit for the two-factor model, other factorial models may also be viable in explaining well-being. Another implication is how counselors conceptualize well-being may be more broadly interpreted relative to the original theoretical underpinnings of the SWBS. As our understanding of QOL indicators have advanced to include other subjective experiences (Bufford et al., 1991), counselor conceptualization of well-being may be changing given evolving developmental, social, cultural, and spiritual dimensions of the human experience. Further, results appear to support Ledbetter, Smith, Fischer et al.'s assertion that "the SWBS may have included items which tap a wide range of behaviors and is not in alignment with

the authors' purpose of unidimensional subscales" (p. 100). The results of this study raise the possibility that constructs measuring other beliefs, behaviors, and affective states may need to factor into our understanding of spiritual well-being.

The complexity and multidimensionality of spirituality and spiritual well-being is evident. It is reasonable to assert that there are inherent challenges with assessments like the SWBS to adequately capture the constructs it seeks to measure. However, there is still utility for counseling professionals in using the SWBS, both with clients and students.

Professional counselors must be mindful of the complexity and nuances associated with client spiritual well-being, and the SWBS, when integrating that dimension into a holistic approach to treatment since clients often bring a unique and personalized perspective to therapy that may not be fully captured with this scale. Further, knowing there are various dimensions that comprise spiritual well-being can not only assist counselors in their conceptualization and practical application of well-being when treating clients, but also provide greater therapeutic clarity when examining how these dimensions shape clients' perceptions of purpose and meaning of life. Using the SWBS may provide an entry point for professionals in understanding these dimensions while being cognizant of the importance of counselor-client dialogue in expanding on and sharpening assessment results.

Counselor educators can look at the present study as evidence for the continued intentional and effective training of counseling students on spirituality and spiritual well-being as a key QOL indicator and necessary component to a holistic approach to counseling. Counselor educators may also gain a better understanding of spiritual well-being and the components that make up this construct. This understanding may enable educators to effectively communicate to students how spirituality/well-being is multidimensional and complex, resulting in reduced theoretical and

practical challenges in assessing. Assessments such as the SWBS are important tools for measuring the continually evolving well-being construct. For students, the SWBS may provide a framework to help clarify their ideas about what spirituality means. It may also help students better conceptualize and integrate their understanding of existential, spiritual, and religious well-being dimensions in their clinical training and work.

Limitations and Future Directions

Although the current study adds to the existing literature regarding the construct validity of the SWBS, several limitations were identified. Using a cross-sectional design limits participants' perception to a single data collection point. Using a self-report measure poses another limitation and can lead to recall bias resulting in potential measurement error. Also, participants who did not identify as spiritual or religious may have been less likely to respond to the survey, possibly resulting in skewed responses. Additionally, it is unknown whether some participants in this study were enrolled in counseling programs in which religious issues and spirituality were emphasized. A curriculum with an emphasis on spirituality and religion would likely influence student responses on the SWBS. Lack of good model fit may be due to concerns with psychometric limitations of the SWBS or statistical artefacts resulting from ceiling effects.

Future research may focus on integrating multi-method designs to improve psychometric data of the SWBS. Investigating the presence of method effects associated with negatively worded items as well as factors that might contribute to it will be beneficial. Continued research is needed to address the inconsistencies in the construct validity of the SWBS as well as the respecified model. Conceptualize well-being across existential, spiritual, or religious domains requires additional research regarding how perceptions may have contributed to the various model fit issues. The findings also reiterate the need for continued research on spiritual and religious

domains (Hall et al., 2004; Miller & Thoresen, 2003; Nelson et al., 2011; Prest et al., 1999; & Sink & Devlin, 2011). Future research should consider conducting other invariance procedures that examine multigroup CFA such as configural invariance, metric invariance, and scalar invariance to test for effects of demographic characteristics in lieu of conducting CFA.

References

- Awang, Z. (2012). *A handbook on SEM* (2nd ed.). https://www.academia.edu/34981094/A_Handbook_on_SEM_Zainudin_Awang_-_Universiti_Sultan_Zainal_Abidin
INTRODUCTION TO STRUCTURAL EQUATION MODELING
- Bufford, R. K., Paloutzian, R. F., Ellison, C. W. (1991). Norms for the Spiritual Well-Being Scale. *Journal of Psychology and Theology*, 19(1), 56-70. <http://journals.biola.edu/jpt?ga=2.12360116.183038728.1499714141-1587409215.1499714141>
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (3rd ed.). Routledge.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Sage.
- Cashwell, C. S., Glossoff, H. L., & Hammond, C. (2010). Spiritual bypass: A preliminary investigation. *Counseling and Values*, 54(2), 162-174. <https://doi.org/10.1002/j.2161-007x.2010.tb00014.x>
- Cashwell, C. S., & Young, J. S. (2004). Spirituality in counselor training: A content analysis of syllabi from introductory spirituality courses. *Counseling and Values*, 48(2), 96-109. <https://doi.org/10.1002/j.2161-007x.2004.tb00237.x>
- Comrey, A., & Lee, H. (1992). *A first course in factor analysis*. Erlbaum.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from our analysis. *Practice Assessment, Research & Evaluation*, 10(7), 1-9. <http://pareonline.net/getvn.asp?v=10&n=7>
- DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Sage.
- DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling: A Multidisciplinary Journal*, 13(3), 440-464. https://doi.org/10.1207/s15328007sem1303_6
- DiStefano, C., & Motl, R. W. (2009). Personality correlates of method effects due to negatively worded items on the Rosenberg Self-Esteem scale. *Personality and Individual Differences*, 46(3), 309-313. <https://doi.org/10.1016/j.paid.2008.10.020>
- Ellison, C. W. (1983) Spiritual well-being: Conceptualization and measurement. *Journal of Psychology and Theology*, 11(4), 330-340. <http://journals.biola.edu/jpt?ga=2.12360116.183038728.1499714141-1587409215.1499714141>
- Fee, J. L., & Ingram, J. A. (2004). Correlation of the holy spirit questionnaire with the Spiritual Well-Being Scale and the Spiritual Assessment Inventory. *Journal of Psychology & Theology*, 32(2), 104-114. http://journals.biola.edu/jpt?_ga=2.12360116.183038728.1499714141-1587409215.1499714141
- Fernander, A., Wilson, J. F., Staton, M., & Leukefeld, C. (2004). An exploratory examination of the Spiritual Well-Being Scale among incarcerated black and white make drug users. *International Journal of Offender Therapy and Comparative Criminology*, 48(4), 403-413. <https://doi.org/10.1177/0306624X04263450>
- Genia, V. (2001). Evaluation of the spiritual well-being scale in a sample of college students. *International Journal for The Psychology of Religion*, 11(1), 25-33. https://doi.org/10.1207/s15327582ijpr1101_03
- Gorsuch, R. L. (1984). The boon and bane of investigating religion. *American Psychologist*, 39(3), 228-236. <https://doi.org/10.1037/0003-066x.39.3.228>

- Hall, S. F., Burkholder, D., & Sterner, W. R. (2014). Examining spirituality and sense of calling in counseling students. *Counseling and Values, 59*(1), 3-16. <https://doi.org/10.1002/j.2161-007X.2014.00038.x>
- Hall, C., Dixon, W., & Mauzey, E. (2004). Spirituality and religion: Implications for counselors. *Journal of Counseling & Development, 82*(4), 504-507. <https://doi.org/10.1002/j.1556-6678.2004.tb00339.x>
- Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Prentice Hall.
- Harrington, D. (2009). *Confirmatory factor analysis*. Oxford.
- Hu, L-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford.
- Koenig, H. G. (2008). Concerns about measuring “spirituality” in research. *The Journal of Nervous and Mental Disease, 196*, 349-355. <https://doi.org/10.1097/nmd.0b013e31816ff796>
- Ledbetter, M. F., Smith, L. A., Fischer, J. D., Vosler-Hunter, W. L., & Chew, G. P. (1991). An evaluation of the construct validity of the Spiritual Well-Being Scale. A confirmatory factor analytic approach. *Journal of Psychology and Theology, 19*(1), 94-102. <https://doi.org/10.1177/009164719101900109>
- Ledbetter, M. F., Smith, L. A., Vosler-Hunter, W. L., & Fischer, J. D. (1991). An evaluation of the research and clinical usefulness of the Spiritual Well-Being Scale. *Journal of Psychology and Theology, 19*(1), 49-55. <https://doi.org/10.1177/009164719101900105>
- MacDonald, D. A., & Friedman, H. L. (2002). Assessment of humanistic, transpersonal, and spiritual constructs: State of the science. *Journal of Humanistic Psychology, 42*(2), 102-125. <https://doi.org/10.1177/002216702237126>
- Marsh, H. W. (1996). Positive and negative self-esteem: A substantively meaningful distinction or artifactor? *Journal of Personality and Social Psychology, 70*(1), 810-819. <https://doi.org/10.1037/0022-3514.70.4.810>
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin, 97*, 562-582. <https://doi.org/10.1037/0033-2909.97.3.562>
- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology, 4*, 169. <https://doi.org/10.3389/fpsyg.2013.00169>
- Miller, G., Fleming, W., & Brown-Anderson, F. (1998). Spiritual Well-Being Scale ethnic differences between Caucasians and African-Americans, *Journal of Psychology and Theology, 26*(4), 358-364. <http://journals.biola.edu/jpt?ga=2.12360116.183038728.1499714141-1587409215.1499714141>
- Miller, W. E., & Thoresen, C. E. (2003). Spirituality, religion, and health. *American Psychologist, 58*(1), 24-35. <https://doi.org/10.1037/0003-066x.58.1.24>
- Moberg, D. O., & Brusek, P. M. (1978). Spiritual well-being: A neglected subject in quality of life research. *Social Indicators Research, 5*, 303-323. <http://dx.doi.org/10.1007/BF00352936>
- Morrison, M., & Borgen, W. A. (2010). How Christian spiritual and religious beliefs help and hinder counselors’ empathy toward clients. *Counseling and Values, 55*(1), 25-45. <https://doi.org/10.1002/j.2161-007x.2010.tb00020.x>
- Morrison, J. Q., Clutter, S. M., Pritchett, E. M., & Demmitt, A. (2009). Perceptions of clients and

- counseling professional regarding spirituality in counseling. *Counseling and Values*, 53(3), 183–194. <https://doi.org/10.1002/j.2161-007x.2009.tb00124.x>
- Myers, J. E., & Williard, K. (2003). Integrating spirituality into counselor preparation: A developmental, wellness approach. *Counseling and Values*, 47(2), 142–155. <https://doi.org/10.1002/j.2161-007x.2003.tb00231.x>
- Nelson, J. A., Kirk, A. M., Ane, P., & Serres, S. A. (2011). Religious and spiritual values and moral commitment in marriage: Untapped resources in couples counseling? *Counseling and Values*, 55(2), 228–246. <https://doi.org/10.1002/j.2161-007x.2011.tb00034.x>
- Paloutzian, R. F., & Ellison, C. W. (1982). Loneliness, spiritual well-being and the quality of life. In L. A. Peplau and D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 224-237). Wiley
- Paloutzian, R. F., & Ellison, C. W. (2009). *Manual for the Spiritual Well-Being Scale: Manual version 1.1*. Life Advance, Inc.
- Prest, L., Russel, R., & D'Souza, H. (1999). Spirituality and religion in training, practice and personal development. *Journal of Family Therapy*, 21(1), 60-77. <https://doi.org/10.1111/1467-6427.00104>
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Routledge.
- Scott, E. L., Agresti, A. A., & Fitchett, G. (1998). Factor analysis of the 'Spiritual Well-Being Scale' and its clinical utility with psychiatric inpatients. *Journal for the Scientific Study of Religion*, 37(2), 314-321. <https://doi.org/10.2307/1387530>
- Sheridan, M., Bullis, R., Adcock, C., Berlin, S., & Miller, P. (1992). Practitioners' personal and professional attitudes and behaviors toward religion and spirituality: Issues for education and practice. *Journal of Social Work Education*, 28(2), 190–204. <https://doi.org/10.1080/10437797.1992.10778772>
- Sink, C. A., & Devlin, J. M. (2011). Student spirituality and school counseling: Issues, opportunities, and challenges. *Counseling and Values*, 55(2), 130–148. <https://doi.org/10.1002/j.2161-007x.2011.tb00027.x>
- Souza, K. Z. (2002). Spirituality in counseling: What do counseling students think about it? *Counseling and Values*, 46(3), 213–217. <https://doi.org/10.1002/j.2161-007x.2002.tb00214.x>
- Tomás, J. M., & Oliver, A. (1999). Rosenberg's Self-Esteem Scale: Two factors or method effects. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 84-98. <https://doi.org/10.1080/10705519909540120>
- Utsey, S. O., Lee, A., Bolden, M. A., & Lanier, Y. (2005). A confirmatory test of the factor validity of scores on the Spiritual Well-Being Scale in a community sample of African Americans. *Journal of Psychology and Theology*, 33(4), 251-257. <https://doi.org/10.1177/009164710503300401>
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with non-normal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56-75). Sage.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Education and Psychology Measurement*, 76(6), 913-934. <https://doi.org/10.1177/0013164413495237>

- You, S., & Yoo, J. E. (2016). Evaluation of the Spiritual Well-Being Scale in a sample of Korean adults. *Journal of Religion and Health*, 55(4), 1289-1299. <https://doi.org/10.1007/s10943-015-0010-0>
- Young, J., Wiggins-Frame, M. & Cashwell, C. (2007). Spirituality and counselor competence: A national survey of American Counseling Association members. *Journal of Counseling and Development*, 85(1), 47-52. <https://doi.org/10.1002/j.1556-6678.2007.tb00443.x>