Implications of a Measurement Invariance Study by Age Cohorts for the Life Satisfaction Survey for Apostolic Women Religious (LSSAWR)

Debra W. Moore[1] & Maria Clara Kreis[2]

[1] Program Evaluation and Research Unit, School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, USA

[2] Department of Spirituality and Pastoral Ministry, Thriving in Ministry, Catholic Theological Union, Chicago, IL, USA

Correspondence: Maria Clara Kreis, Department of Spirituality and Pastoral Ministry, Thriving in Ministry, Catholic Theological Union, Chicago, IL, USA.

Abstract

A measurement invariance study was performed on the Life Satisfaction Survey for Apostolic Women Religious (LSSAWR) across age cohorts to support its continued use to assess satisfaction with religious life across an individual sister's life span, and to conduct intergenerational comparisons within and across congregations worldwide. Unfortunately, measurement invariance (MI) is often assumed rather than tested but is important to determine when comparisons are conducted across groups. Establishing MI produces confidence that the differences observed are a result of real differences between groups rather than a result of group membership. In general, the current study provides evidence that the LSSAWR is MI for life satisfaction across the Silent, Baby Boomer, and Generation X cohorts and should be robust to many types of analyses. Therefore, the continued use of the LSSAWR to provide feedback to individual Sisters and congregations of women religious regarding commitment to religious life and overall life satisfaction is supported. The most notable result was two of the five dimensions of the scale were statistically indistinguishable for the Silent generation, but not for the Baby Boomer or Generation X cohorts. This article discusses the importance of measurement invariance studies and implications for instruments used across the life span with items that could be age sensitive.

Keywords: life satisfaction, measurement invariance, age cohorts, developmental life span, religious life, apostolic women religious

1. Introduction

The world has undergone drastic changes due to varied technological advancements of modernity. These changes continue to affect the well-being of people within our current postmodern time. The assessment of life satisfaction across the human life span development is one way to assess how societal, political, religious, and ecological changes might influence satisfaction levels of individuals and the general population. For example, the field of psychology can control for demographic variables (e.g., age, generation cohorts, ethnicity, gender, etc.) when assessing life satisfaction levels across the general population and within diverse subgroups as determined by someone's lifestyle such as partnership, marriage, singlehood, or consecrated life (Diener, Emmons, Larsen, & Griffin, 1985; Kreis, 2012; Schumm, Nichols, Sheetman, & Grigsby, 1983). An important psychometric property of any assessment tool used to compare scores across subgroups within the sample is measurement invariance (MI). The assessment of measurement invariance is fundamental for instruments and tools used in research and professional practice within the field of psychology to ensure impartial comparative investigations so that observed differences can be attributed to true differences between groups and not a statistical artifact due to group membership (Büchi, 2016; Byrne, Shavelson & Muthén, 1989). Surprisingly, despite its importance there are many assessment instruments that are being used to conduct research nationally and internationally that until recently have not undergone an assessment of MI (Sischka, Costa, Steffgen, & Schmidt, 2020; Yap, Donnellan, Schwartz, Kim, Castillo, Zamboanga, Weisskirch, Lee, Park, Whitbourne, & Vazsonyi, 2014).

Furthermore, when MI was assessed, many of the instruments being used for gender, ethnic or cross cultural research indicated metric and configural invariance, but were less likely or only partially able to demonstrate scalar invariance across ethnic or cultural groups (Lin, Chen, Tan, Yang, & Chi, 2021; Sischka et al., 2020; Sorrel, García, Aluja, Rolland, Rossier, Roskam, & Abad, 2021; Yap et al., 2014). An example is the MI research conducted on the widely used 'Satisfaction with Life Scale' by Diener et al. (1985), where the cross-cultural research team found metric and configural but not scalar invariance across samples composed from 29 countries worldwide (Jang, Kim, Cao, Allen, Cooper, Lapierre, Driscoll, Sanchez, Spector, Poelmans, Abarca, Alexandrova, Antoniou, Beham, Brough, Carikci, Ferreico, Fraile, Guerts, Kinnunen, Lu, Lu, Moreno-Velázquez, Pagon, Pitariu, Salamtov, Siu, Shima, Schulmeyer, Tillemann, Widerszal-Bazyl, & Woo, 2017). This highlights the importance of MI on instruments that are also being used for cross-cultural research. In fact, Razmus, Razmus, Tylka, Jović, Jović, and Namatame (2020) emphasized the importance of conducting a MI assessment before any comparisons are done across cultural groups. Therefore, the focus of this article will be on assessing the MI of the LSSAWR across age cohorts prior to future cross-cultural group comparisons.

1.1 Exploration of the Problem

Roman Catholic Women Religious are an intergenerational group of women who have joined the consecrated life within their respective congregations worldwide. Historically, women religious are known for their spiritual and communal life. This lifestyle inspires and strengthens their ability to commit their lives to altruistic service to those most disadvantaged in society and the world (Ebaugh, Lorence, & Saltzman Chafetz, 1996). Nevertheless, their lives have been affected by an interplay of external and internal dynamics due to sociological, political, and religious changes since the 1950s (Kreis, 2010; Kreis, 2012; Wittberg, 1994). Currently, women religious are undergoing major transitions as they seek to redefine their identity, purpose, and mission nationally and internationally. The use of an instrument known as the Life Satisfaction Scale for Apostolic Women Religious (LSSAWR) has shown positive results. In fact, women religious who have used the LSSAWR reported that it has offered life-giving directions as they individually and communally reflected on their life, and strategically planned and transitioned into their preferred future (Kreis, 2019).

The LSSAWR is designed to offer immediate individual feedback online to women religious as well as congregational feedback through anonymous aggregated data provided in Congregational Reports (LSSAWR website). Currently, there are four generational cohorts in religious life and the age range of women religious within a congregation can span from 18 to over 100 years of age. In applying an interdisciplinary research approach between developmental psychology (Erickson, 1959) and sociology's generation theory (Strauss & Howe, 1991), it is therefore necessary to assess whether the LSSAWR can be used to a) evaluate satisfaction with religious life across an individual sister's life span, and b) conduct intergenerational comparisons within and across congregations worldwide.

The test-retest reliability and factor structure of the LSSAWR were established by Kreis (2010; 2012) and Kreis, Crammond, and Reynolds (2018). However, the measurement invariance of the LSSAWR among different subgroups, specifically age cohorts, has yet to be investigated. Consequently, it is imperative to evaluate the MI of the LSSAWR to support its continued use in research across generations of women religious congregations worldwide (LSSAWR website).

Hence, the crucial research question guiding this study is whether the LSSAWR is measurement invariant across generational cohorts such that it can be used to research life satisfaction across the life span and among different age cohorts.

1.2 The LSSAWR

The Life Satisfaction Scale for Apostolic Women Religious (LSSAWR) is the first instrument designed to measure satisfaction with religious life among Catholic Sisters (Kreis, 2010; Kreis, 2012). The instrument consists of 50 statements concerning the life of a Roman Catholic apostolic woman religious. Psychometric examination of the first wave of data collected revealed 5 underlying dimensions associated with life satisfaction as measured by the LSSAWR: Congregational Character, Individual Well-Being, Membership Viability, Holistic Growth and Commitment, and Inter-Relationships. Sisters taking the instrument rate their satisfaction with the statements on a 5-point Likert scale from 'Very dissatisfied' to 'Very satisfied.' Scores for each item within each dimension are summed to provide information about a Sister's satisfaction with the referenced aspect of religious life. The scores from the 5 dimensions are combined to calculate a total score that provides a general satisfaction index with religious life (Kreis et al., 2018).

1.3 LSSAWR Research and Congregational Reports

The LSSAWR is currently used to provide feedback to individual Sisters and congregations of women religious regarding commitment to religious life and overall life satisfaction (Kreis, 2010; Kreis & Diaz, 2021). There have been 3 waves of data collection for the survey. Wave 1 was collected in 2008-2009 and used for the original factor analysis and psychometric evaluation. Wave 2 was collected in 2016-2020 and included the development of the LSSAWR Manual (Kreis et al., 2018). This wave of data collection expanded efforts to recruit Sisters internationally from different generational cohorts, and to offer a Spanish and a German version of the instrument and manuals (Kreis, Crammond, & Reynolds, 2019; a, b). Wave 3 was collected in 2021 and featured gender inclusive language, instead of her or she, in an effort to expand the use of the instrument to include men religious. Congregational Reports are provided by request that explain the instrument results from women religious congregation for each of the 5 dimensions and overall. Based on sufficient numbers within each age cohort, Congregational Reports may include the presentation of generational differences in the form of average item score by dimension. The generation cohorts in (Kreis, 2010; 2012) follow the age cohort categories from the oldest to the youngest as established by Strauss and Howe (1991). Currently, the data collection of the LSSAWR includes women religious worldwide who belong to the Silent Generation (1925-1942), Baby Boom Generation (1943-1960), Generation X (1961-1981), and Millennial Generation (1982-2004) age cohorts.

An important consideration in comparing observed differences among groups is assuring the same underlying construct of interest is measured in the same way across groups (Büchi, 2016; Byrne et al., 1989). Changes in the meaning of words over time and differences in individual-level understanding of survey statements can impact the accuracy of interpretations of any observed differences in scores (Tucker, Ozer, Lyubomirsky, & Boehm, 2006). Establishing factorial invariance or measurement equivalence is important in demonstrating that observed differences reflect true differences in the underlying construct being measured as opposed to differences that can be attributed to group membership (Cheung & Lau, 2012; Putnick & Bornstein, 2016). However, MI is often assumed rather than actually tested, and ignoring it can lead to biased results (Guenole & Brown, 2014; Wu, Li, & Zumbo, 2007). Regarding the application and inconsistent reporting of MI, there appears to be a lack of understanding and/or guidance in its use and its impact when using instruments and tests in research and the applied field (Putnick & Bornstein, 2016; Schweig, 2014).

1.4 Measurement Invariance

MI is most frequently tested within a multigroup confirmatory factor analysis (MG-CFA) framework (Vandenburg & Lance, 2000; Asparouhov & Muthén, 2014; Putnick & Bornstein, 2016). Typically, three levels of invariance are established to provide evidence that an instrument is measuring the same construct in the same way across groups: configural, metric, and scalar. A fourth level of invariance, residual invariance, also exists in literature (Putnick & Bornstein, 2016). However, since residual invariance is not required to test mean differences, it is rarely tested in practice (Meredith, 1993; Putnick & Bornstein, 2016). Within the framework of MG-CFA, configural invariance establishes that the number of factors and items associated with each factor (factor structure) are similar across groups (Cheung & Lau, 2012). This level of invariance is often called weak invariance because it represents what is considered the lowest level of MI (Meredith, 1993). Once configural invariance is established, metric invariance is examined. Metric invariance establishes whether the factor to item relationships (factor loadings) are the same across groups. This is often referred to as strong invariance (Tucker et al., 2006). If both configural and metric invariance are established, finally, scalar invariance is examined. Scalar invariance establishes that the groups being tested have approximately the same intercepts or thresholds (Meredith, 1993; Cheung & Lau, 2012; Putnick & Bornstein, 2016). From a practical perspective, scalar invariance answers questions such as, do responses from the Baby Boomer generation have similar interpretations for all scale points, including the zero point, as the other generational cohorts? This concrete application of scalar invariance provides a practical reason for the finding from research that shows scalar non-invariance impacts the comparison of latent variable means to a larger extent than metric non-invariance (Steinmetz, 2013; Schmitt, Golubovich, & Leong, 2011).

Therefore, when an instrument fails to meet any level of MI, typical remedies such as identifying and removing problematic items or establishing partial invariance are undertaken (Asparouhov & Muthén, 2014; Putnick & Bornstein, 2016). Unfortunately, as Asparouhov and Muthén point out, removing problematic items may mean substantial change to the original model specification that can lead to model misspecification which, in turn, defeats the purpose of the MI study. Alternatively, researchers often establish partial invariance (Asparouhov & Muthén, 2014; Putnick & Bornstein, 2016; Jung & Yoon, 2016). Establishing partial invariance involves systematically removing constraints and retesting the model for MI (Jung & Yoon, 2016). However, no consensus exists about the impact partial invariance has on comparative studies or what methodology represents

best practice for establishing non-invariance (Jung & Yoon, 2016). Finally, in some cases, researchers may argue that the lack of invariance represents a meaningful difference that could be informative on its own (Cheung & Rensvold, 1999; Putnick & Bornstein, 2016).

Still, in reality, MI across all three levels is difficult to achieve and failure to achieve full invariance, and scalar invariance in particular, is not uncommon (Vandenburg & Lance, 2000; Asparouhov & Muthén, 2014; Putnick & Bornstein, 2016). Examination of invariance across the levels generally requires comparing a series of nested models with increasingly stricter equality constraints through the lens of specific model fit indices (Widamen & Reise, 1997; Vandenberg & Lance, 2000). The typical model fit index for comparing constrained and unconstrained, or increasingly constrained nested models, is the chi-square difference test (Bollen, 1989; Putnick & Bornstein, 2016). However, this test is sensitive to sample size and even small differences in sample size can lead to rejection of MI (Brannick, 1995; Chen, 2007; Cheung & Rensvold, 2002; French & Finch, 2006; Putnick & Bornstein, 2016). As one response to this, it was recommended to use change in the Comparative Fit Index (CFI) between nested models to evaluate MI (Cheung & Rensvold, 2002; Kline, 2015). However, as Cheung and Rensvold (2002) point out, there is no distributional assumptions for CFI and, therefore, no significance test is available. This opens the practice of comparing fit indices to criticism (Putnick & Bornstein, 2016). Another popular model fit index is the Root Mean Square Error of Approximation (RMSEA) because it adjusts for model complexity and is less sensitive to sample size (Putnick & Bronstein, 2016). However, recent research has determined that the same cutoff for any of the fit indices may not be applicable to every situation due to differences in the number of factors in the model, number of items on the scale, measurement level of the item (e.g., ordinal), and number of groups being examined among other factors (Svetina, Rutkowski, & Rutkowski, 2020; Rutkowski & Svetina, 2014; Fan & Sivo, 2009; Kenny, Kanistan, & McCoach, 2015). In fact, Chen (2007) found that these typical fit indices are impacted by unequal sample sizes between groups being compared and whether or not real differences exist in factor loadings and item variances and intercepts. Additionally, Rutkowski and Svetina (2014) found that as the number of groups increases, CFI becomes more sensitive and the RMSEA becomes less sensitive to non-invariance. Fan and Sivo (2009) found that both CFI and RMSEA become less sensitive as the model grows in size (e.g., more factors, more items). These studies have led many researchers to recommend conditional cutoffs depending on model complexity and other parameters of the model (e.g., number of groups, sample size, unequal group sizes; Fan & Sivo, 2009; Kenny et al., 2015). Therefore, the authors focused their MI evaluation on the application of change in the CFI and RMSEA using values from recent studies that most closely matched the parameters of the current study.

2. Method

In order to more deeply explore the psychometric properties of the LSSAWR and its ability to support robust comparison of age cohort differences in life satisfaction among women religious, the current study examined MI across generational cohorts for the LSSAWR as established by Strauss and Howe (1991).

2.1 Sample

The sample for the study consisted of responses from Wave 2 of LSSAWR data collection between 2016 and 2020 (Kreis, 2020; Kreis & Crammond, 2019). This wave of data collection was chosen because it had the highest representation of age cohorts and was international in nature. A generational cohort variable, consistent with Silent (1925-1942), Baby Boomer (1943-1960), Generation X (1961-1981), and Millennial (1982-2004) was created. Additionally, this key variable was given a numeric code (e.g., Silent generation coded as 1, etc.) to facilitate the MI investigation.

The sample size for the analysis was 1,890 with the demographic frequencies presented in Table 1. The proportion of the sample from the Silent generation and the Baby Boomer generation were approximately equal (roughly 36% each) with the Generation X cohort (20%), and Millennial cohort (7%) comprising considerably smaller proportions of the sample. Unsurprisingly, given their age, Millennials were the least represented cohort in the sample. Most respondents took the survey in English and, anecdotally, other languages were more likely to be clustered by congregation, especially Spanish. These respondents did offer their voluntary and anonymous participation in taking the LSSAWR online as they replied to the invitation expressed by their leadership teams, who were collaborating with the LSSAWR research team in receiving a Congregational Report based on their congregations' LSSAWR scores. A large majority of the respondents (76%) had at least a bachelor's degree with just over half of the sample (57%) having a master's degree or higher educational attainment. Most respondents were of European ancestry with small representation each from Asian, African, and Hispanic ancestry.

Table 1. Demographic frequencies for study sample

Variable	Levels	n	%
Generational Cohort	Silent	689	36.5%
	Baby Boomer	668	35.3%
	Generation X	371	19.6%
	Millennial	136	7.2%
Race/Ethnicity	African American	12	<1%
	Asian American	27	1.4%
	European American	1005	53.2%
	Native American	20	1.1%
	Hispanic American	174	9.3%
	Other or more than 1 category (US born)	40	2.1%
	Other or more than 1 category (not US born)	231	12.3%
	German	207	11%
	European	165	8.8%
Educational	Some HS	91	4.8%
Attainment	HS diploma	200	10.6%
	Some college credits	107	5.7%
	Associate degree	58	3.1%
	Bachelor's degree	351	18.6%
	Master's degree	598	31.7%
	Post-Master's	323	17.1%
	Doctoral	128	6.8%
	Post doc	31	1.6%
Survey Language	English	1302	68.9%
	German	373	19.7%
	Spanish	215	11.4%

2.2 Procedure

A MG-CFA for generational cohort was performed in MPlus v.7.2 (Muthén & Muthén, 2012). Results of the MG-CFA were evaluated using change in fit indices (Svetina et al., 2020). Based on psychometric work already completed, the model specified for the MI study was the same 5-factor model outlined in the LSSAWR Manual. Additionally, a crosstab was performed to assess the completeness of data by generational cohort. Results of the crosstab revealed a large number of items (n = 16; 32% of all items) for which no respondents in the Millennial cohort endorsed the lowest response category (very dissatisfied). This, in conjunction with the large proportional difference in sample size between the Millennial generation and the Silent generation (n = 689; 6 times larger) and the number of parameters to be estimated in the analysis compared to observations (n = 136) caused concern about including the Millennial generation in the MI study. Ultimately, they were removed, and the study was conducted among the three remaining generational cohorts; Silent, Baby Boomer, and Generation X. In addition, the crosstab revealed that either the Boomer cohort, Generation X cohort, or both cohorts did not have anyone endorse the lowest response category (very dissatisfied) for 5 items (Items 15, 31, 32, 33, 34, 50; 12% of all items). Item 15 asked about professional relationships and was associated with the Inter-Relationship dimension. Items 31 – 33 were associated with the Holistic Growth and Commitment dimension and asked about personal growth, spiritual growth, and relationship with God. Item 34 was associated with the Congregational Character dimension and asked about congregational prayer and rituals. Item 50 asked the respondent to rate their overall satisfaction with commitment to their congregation and was associated with the Individual Well-Being dimension. These 5 items were removed from the MI analyses so the model would estimate without flagging

these items in the technical indices. In general, response category endorsement was highly negatively skewed for almost all items.

However, even with these modifications, the model testing configural invariance would not estimate due to misspecification. Technical outputs were requested since the metrics associated with these tests show item statistics and correlations between items and factors that highlight problems with the models (Asparouhov & Muthén, 2014). Examination of these technical outputs showed that the Individual Well-Being dimension (with Item 50 excluded) and the Inter-Relationships dimension (with Item 15 excluded) were statistically indistinguishable for the Silent generation, but not for the Baby Boomer and Generation X cohorts.

Therefore, a second MG-CFA by generational cohort was performed in MPlus v.7.2 (Muthén & Muthén, 2012) collapsing the two identified factors into a single factor. This resulted in a 4-dimensional model being tested for the second MI study. In preparation for this second MI study, a CFA using all responses was performed and fit indices were compared to examine the suitability of the collapsed factor structure (Table 2).

Table 2. Model fit statistics for the 5-factor and 4-factor structure

Model Fit Statistic	5-factor	4-factor
RMSEA	0.073	0.081
CFI	0.861	0.832
TLI	0.853	0.823
Correlations among factors	0.548 – 0.780	0.619 – 0.830

After examining fit statistics, the MI study with the Individual Well-Being dimension and Inter-Relationships dimension combined was completed. For clarity, it is noted that the second analyses still excluded the Millennial cohort and the 5 items previously identified. Due to the large sample size, results of the MG-CFA did not rely solely on the X^2 statistic and were evaluated using change in model fit indices.

3. Results

Both MG-CFA studies used ordinal scale items (Likert items) associated with 4 factors (2 of the 5 factors were combined) to compare 3 generational cohorts with unequal sample sizes ranging from approximately 371 to 689. Results are presented in Table 3. All model estimations terminated normally.

Table 3. Fit statistics and calculations for second MI study

Fit Statistic	Configural Model	Metric Model	Scalar Model	Metric-Configural	Scalar-Metric
RMSEA	0.072	0.070	0.066		
CFI	0.881	0.887	0.885		
X^2 diff test*				612.0	868.9
df				80	344
p				< 0.001	< 0.001
Δ RMSEA				-0.002	-0.004
Δ CFI				0.006	-0.002

NOTE: run using diff test commands for nested models in Mplus

3.1 Determining Appropriate Cutoff Values

Asparouhov and Muthén, (2014) demonstrated the complexity of comparisons across more than 2 groups in establishing MI. Recently, other researchers (e.g., Svetina et al., 2020; Rutkowski & Svetina, 2017; Svetina & Rutkowski, 2017) have demonstrated that the number of factors being investigated, sample size, unequal group sizes, and the use of categorical variables add additional complexity in interpreting MI study results. This has led to a revival of alternatives to the traditional MG-CFA methodology of establishing MI and the use of change in fit statistics, supported by Monte Carlo simulation studies, to assess the invariance of measurement instruments (Svetina et al., 2020). Differences in sample size, number of groups, number of factors and items, level of measurement (e.g., ordinal) across these studies have resulted in different recommendations for change in model fit criteria (Putnick & Bornstein, 2016).

Studies using normal models, as opposed to categorical models, suggest ΔCFI ≥ -0.005 and ΔRMSEA ≤ 0.01 (Chen, 2007); ΔCFI ≤ -0.10 (French & Finch, 2006); ΔCFI ≥ -0.01 (Cheung & Rensvold, 2002), or ΔCFI < 0.01

(Kim, Cao, Wang & Nguyen, 2017), or ΔCFI $\geq$ -0.02 and -0.01 and ΔRMSEA $\leq$ 0.03 and 0.01, for metric and scalar invariance, respectively (Rutkowski & Svetina, 2014). More appropriately, two studies used ordinal data, a MG-CFA design, and/or more than 1 factor and sample sizes similar to the current study. Rutkowski and Svetina (2017) suggested a ΔRMSEA $\leq$ 0.05 with a significant change in chi-square value and a ΔCFI $\geq$ -0.004. The suggested ΔRMSEA value was slightly smaller (0.01) when considering scalar invariance. This was in the context of a MG-CFA with 10 or 20 groups, a unidimensional scale, and samples sizes ranging from 600 to 6,000. Finally, within an MG-CFA study with 10 or 20 groups ranging in size from 750 to 6,000 and 2 or 5 dimensions, Svetina and Rutkowski (2017) suggested similar values as in the Rutkowski and Svetina (2017) study, except for a ΔCFI $\geq$ -0.002. Applying the suggested cutoffs based on studies with parameters as similar as possible to this study, seemed the most appropriate method to determine MI for comparing age cohort scores on the LSSAWR.

3.2 Interpretation of Results

The design of the Svetina and Rutkowski (2017) study most closely matched the parameters of the current study. However, a conservative approach of applying the cutoff criteria from both studies was applied to evaluate the MI of the LSSAWR. Table 4 compares the results of the current study with the suggested cutoff values from these studies.

Table 4. Current study results compared to cutoffs from relevant studies

Study	Model Fit Statistic	Recommended Cutoff	Metric	Scalar
Rutkowski and Svetina (2017)	Δ RMSEA	$\leq$ 0.05 with sign. X^2	√	√
	Δ CFI	$\geq$ -0.004	√	√
Svetina and Rutkowski (2017)	Δ RMSEA	$\leq$ 0.01 with sign. X^2	√	√
	Δ CFI	$\geq$ -0.002	√	√

Comparing the current study's ΔRMSEA and ΔCFI values to the above criteria, the LSSAWR meets the criteria suggested for metric invariance and scalar invariance recommended by both studies. It is also interesting to note that the ΔRMSEA and ΔCFI values from the current study also meet all but one criterion for metric invariance and all but two criteria for scalar invariance recommended by the additional 5 studies conducted using normal models. These results strongly suggest that the LSSAWR meets the criteria for configural, metric, and scalar invariance by age cohort for the 4-dimensional structure used in the study, especially based on cutoffs determined in studies that most closely mimic the parameters of the current study.

4. Discussion

The current MI study provided evidence that the LSSAWR is measurement invariant at the configural, metric, and scalar levels for overall life satisfaction and for 3 of the 5 dimensions across the Silent, Baby Boomer, and Generation X cohorts. The study strongly supports the continued use of providing feedback to individual sisters across their life span concerning satisfaction with religious life. The study also supports the reporting and analysis of generational differences across the Silent generation, Baby Boomer, and Generation X cohorts concerning Overall Life Satisfaction. (Kreis et al., 2018; Kreis et al., 2019; a, b) In addition, the reporting and analysis of generational differences across Congregational Character, Membership Viability, and Holistic Growth and Commitment is also supported. Reporting or comparing the Individual Well-Being score and the Inter-Relationships score for the Baby Boomer and Generation X cohorts is supported as well. Caution is warranted, however, when comparing the Individual Well-Being score to the Inter-Relationships score across the generational cohorts when the eldest age cohort (Silent generation) is included, as these dimensions do not appear to be distinct within members of the eldest age cohort (Silent generation). Certainly, creating a combined score across both these dimensions and comparing scores across age cohorts is supported as is the use of a combined dimensional score as part of an analysis using variables obtained from other sources. Therefore, the LSSAWR is robust to many types of analyses using age cohorts. As stated earlier, more data and analyses are needed before comparisons and analyses using the Millennial generation can be made with confidence.

Current research affirms that sample size, model complexity (number of factors and number of items loading on each factor), and unequal sample sizes among comparison groups impact a MI study (Svetina et al., 2020) Therefore, the Millennial generation (born 1982-2004) was not included in the analysis. The large proportional difference between the small sample size of this group compared to the largest group (n = 136 versus n = 689), the large proportion of items (32%) for which no Millennial endorsed the lowest response choice, and the number of parameters needing estimation compared to the number of observations, support the removal of this

age cohort at this time. However, the small sample size of this group is not unexpected given that many in this cohort are not yet 18 years old and are most likely not in any commitment yet. Therefore, it is important to continue data collection with the Millennial cohort and repeat the analyses outlined in this paper. A larger proportion of Millennials will provide more information and/or additional confirmation of the current results and interpretations.

The most notable result was that two of the five dimensions, Individual Well-Being, and Inter-Relationships, were statistically indistinguishable for the oldest of the three generational cohorts, the Silent generation, but not for the Baby Boomer or Generation X cohorts. Fit statistics for the 5-dimension CFA performed on the combined sample were in the acceptable range and slightly better than the fit statistics for the 4-dimension CFA performed, confirming the original 5-dimensional structure of the instrument established in 2016. Regardless, respondents from the Silent generation appear to be interpreting items associated with the Individual Well-Being dimension and the Inter-Relationships dimension as being more similar or more highly related than the Baby Boomer or Generation X cohort. This result highlights the importance of examining measurement invariance for groups within the sample in addition to confirming the underlying factor structure, especially if additional analyses are desired in the future.

While it is not unprecedented to find instruments that meet the criteria for configural, metric, and scalar invariance, it is also not unusual for instruments to fail to meet the criterion for scalar invariance (Lin et al., 2021; Sischka et al., 2020; Sorrel et al., 2021; Svetina et al., 2020; Vandenburg & Lance, 2000; Yap et al., 2014). The homogeneous nature of this population and singular commitment of their lives to altruistic service, however, allows for the 'control' of variables (e.g., marital status, children, life goals) that could impact the measurement invariance of life satisfaction instruments across a more heterogeneous sample. In fact, conducting an MI study on this homogeneous population has highlighted a potential area for additional research and intervention to improve life satisfaction across the life span.

Items associated with the Individual Well-Being dimension focus on validation and support received, connectedness to their congregation, the quality of relationships among the Sisters in their congregation, and their leadership, ministry, and overall contributions to the congregation (Kreis et al., 2018). Items associated with the Inter-Relationships dimension focus on maintaining professional, family, and friend relationships outside of their congregation, close relationships within their congregation, and inclusion of others and interactions with people in their own age group (Kreis et al., 2018). Given the natural decrease in personal and professional relationships due to advanced age, death of family and friends, and deteriorating health among sisters of the Silent generation (currently between the ages of 79-96), it is not surprising that these types of relationships are less influential in their overall life satisfaction. As time progresses, these relationships no longer exist or have been marginalized due to other factors. Furthermore, elder women religious who experience a decrease in one or more areas of their overall well-being (biological, psychosocial, spiritual), tend to withdraw from active ministry and active involvement in community life as they graciously move into a fulltime prayer ministry (Kreis & Diaz, 2021). Therefore, items that reflect active ministry involvement and contributions to community life could be less influential in their lives. It is also noteworthy, that the data collection for this instrument has taken place over more than a decade. In fact, Wave 1 (original and national sample) and Wave 2 (international sample) had different congregations of women religious participate. Participating congregations in Wave 1 were all located in the USA, while individuals and congregations of women religious who participated in Wave 2 were from many countries and represented every continent of the world. Furthermore, members of all age cohorts in the Wave 2 data collection were at least 10 years older than those participating in the national data collection in 2008-2009. One can assume that at least a proportion of the 2009 Silent generation (national sample) were more actively involved in living their commitment as women religious and ministering to the world (Kreis & Diaz, 2021). These observations tend to imply, that the older in age and the more compromised in their well-being women religious of the eldest generation cohort are, the more likely it is that their LSSAWR results will indicate a less distinctive impact between these two factors (Individual Well-Being and Inter-Relationships) on their lives.

4.1 Limitations

A few limitations exist in the current study. First, the Wave 2 cohort had a large number of congregational leaders from the Silent generation. Some of these respondents, through the qualitative comments, expressed confusion about how to answer some of the relationship items pertaining to leadership. Position within the congregation, however, is not a demographic question asked, so the exact proportion of respondents in a leadership position is unknown. It is of note that leaders could interpret items associated with the Individual Well-Being dimension and the Inter-Relationships dimension differently because of their leadership position and

its influence on their relationships with others in their respective congregations. Instructions were provided with the survey concerning this, but based on information provided in the qualitative comments, they were not always understood or considered.

Second, Millennials did not endorse the lowest category for a large proportion of items. More importantly, the small sample size for this age cohort posed two specific problems for the MI model estimation. Additionally, there was a large proportional difference between the size of this cohort and the largest age cohort. Unequal sample sizes in groups being compared during an MI study has been shown to impact results (Chen, 2007). It is also important that the number of observations (sample size) needs to be considerably larger than the number of parameters needed for the analyses, otherwise the model will not estimate correctly. Hence, the Millennial cohort had to be excluded from the study. Whether the lack of endorsement across all response categories was due to higher satisfaction levels or the comparatively smaller sample size of this international sample or some combination of both is unknown. Responses from this cohort were used to complete the CFAs that confirmed the 5-dimensional and 4-dimensional models.

Third, there are other ways in which the sample could be divided into groups; survey language, ethnicity, gender, educational attainment, and use of gender inclusive language. Each of these different categorizations could interact in different ways to influence this MI analysis. In the future, it will be important to examine differences in these additional categories, where and when sample size allows, to determine the appropriateness of cross-category research with these variables.

Finally, three somewhat problematic items were identified during the examination of the technical indices. These items did not interfere with completing the MI study but were flagged for continued consideration and investigation. Items 17 and 18 had the lowest factor loadings in both Wave 1 and Wave 2 data (Kreis, 2010; 2012; Kreis et al., 2018; 2019; a, b.). These two items measure opinions about relationships with the Roman Catholic Church hierarchy. Additionally, Item 27, especially in correlation to Items 26 and 28, was not performing well. This set of items (26, 27, and 28) measured opinions about leadership with Item 27 being specific to administrative skills. As noted earlier, several respondents were administrators and expressed in the open-ended comments that they had difficulty in answering this question, even though the survey instructions specified how to respond to this question.

4.2 Summary

Normally, caution would be advised when opting to make comparisons among all 4 generational cohorts because Millennials were not included in the study. However, it is noteworthy that it is the oldest generation in the sample for which differences were directly observed, and that those differences involved a deterioration of distinctness among dimensions. In fact, the results of this study suggest caution be applied when comparing generational scores on the Individual Well-Being and Inter-Relationships dimensions across generational cohorts that include the Silent generation, since the dimensions appear to remain more distinct for the younger generational cohorts of Roman Catholic apostolic women religious. However, in the future, the caution pertaining to the current eldest generation cohort (Silent) might also apply to the Baby Boomer cohort as they become older in age, experience changes within their overall well-being, and decide to withdraw from active ministry and community involvement. Therefore, while the proposed reasoning for the observed differences in the Silent generation are not confirmed, it may be that caution is warranted when making comparisons between the Individual Well-Being dimension to the Inter-Relationships dimension on the LSSAWR using ANY generational cohort as they become the oldest generation in the sample. This argues that these two dimensions may provide separate information about influences on life satisfaction and add value to the interpretation of scores for the younger age cohorts but not the oldest cohort. Future research can be directed toward determining if the LSSAWR can assist with determining the specific age within the oldest age cohort, when reached, which is the turning point for the change.

While additional data collection and analyses will confirm or refute the claims outlined above, the continued use of the LSSAWR to provide feedback to individual Sisters and congregations of women religious regarding commitment to religious life and overall life satisfaction is supported by the results of this measurement invariance study. Additionally, it is recommended that the Individual Well-Being and Inter-Relationships dimensions be retained as separate dimensions for reporting, tracking, and research purposes, at least among the younger generational cohorts when using the LSSAWR. In the larger context of age cohort comparisons, any instrument that takes a developmental approach across the life span and contains items that could be age sensitive, such as pre-retirement work, professional, and family relationships, should be evaluated for measurement invariance especially when conducting comparisons across age cohorts.

Acknowledgments

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495-508. https://doi.org/10.1080/10705511.2014.919210

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley. https://doi.org/10.1002/9781118619179

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*, 201-213. https://doi.org/10.1002/job.4030160303

Büchi, M. (2016). Measurement invariance in comparative Internet use research. *Studies in Communication Sciences, 16*(1), 61-69. https://doi.org/10.1016/j.scoms.2016.03.003

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 3*, 456-466. https://doi.org/10.1037/0033-2909.105.3.456

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

Cheung, G. W., & Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organizational Research Methods, 15*(2), 167-198. https://doi.org/10.1177/1094428111421987

Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management, 25*, 1-27. https://doi.org/10.1016/S0149-2063(99)80001-4

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality and Social Psychology, 49*, 71-75. https://doi.org/10.1207/s15327752jpa4901_13

Ebaugh, H. R., Lorence, J., & Saltzman Chafetz, J. (1996). The growth and decline of the population of Catholic nuns cross-nationally, 1960-1990: A case of secularization as social structural change. *Journal for the Scientific Study of Religion, 35*, 171-183. https://doi.org/10.2307/1387084

Erikson, E. H. (1959). *Identity and the Life Cycle*. New York: International Universities Press.

Fan, X., & Sivo, S. A. (2009). Using ΔGoodness-of-Fit indexes in assessing mean structure invariance. *Structural Equation Modeling, 16*, 54-69. https://doi.org/10.1080/10705510802561311

French, B. F., & Finch, W. H. (2006). Confirmatory factor analytic procedures for the determination of measurement invariance. *Structural Equation Modeling, 13*, 378-402. https://doi.org/10.1207/s15328007sem1303_3

Guenole, N., & Brown, A. (2014). The consequences of ignoring measurement invariance for path coefficients in structural equation models. *Frontiers in Psychology, 5*, 980. https://doi.org/10.3389/fpsyg.2014.00980

Jang, S., Kim, E. S., Cao, C., Allen, T. D., Cooper, C. L., Lapierre, L. M., Driscoll, M. P., Sanchez, J. I., Spector, P. E., Poelmans, S. A. Y., Abarca, N., Alexandrova, M., Antoniou, A. S., Beham, B., Brough, P., Carikci, I., Ferreico, P., Fraile, G., Guerts, S., Kinnunen, U., Lu, C., Lu, L., Moreno-Velázquez, I. F., Pagon, M., Pitariu, H., Salamtov, V., Siu, O., Shima, S., Schulmeyer, M. K., Tillemann, K., Widerszal-Bazyl, M., & Woo, J. M. (2017). Measurement invariance of the Satisfaction with Life Scale across 26 countries. *Journal of Cross-Cultural Psychology, 48*(4), 560-576. https://doi.org/10.1177/0022022117697844

Jung, E., & Yoon, M. (2016). Comparisons of three empirical methods for partial factorial invariance: forward, backward, and factor-ratio tests. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4). https://doi.org/10.1080/10705511.2015.1138092

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods and Research, 44*, 486-507. https://doi.org/10.1177/0049124114543236

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling, 24*, 524-544. https://doi.org/10.1080/10705511.2017.1304822

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York: Guilford Press.

Kreis, M. C. (2010). *Assessment of Life Satisfaction in Apostolic Women Religious: The Development of a New Instrument* (Dissertations, Paper 46).

Kreis, M. C. (2012). The development of a new instrument for the assessment of life satisfaction in apostolic women religious. *Journal of Spirituality in Mental Health, 14*(2). Copyright © in 2018 in English, Spanish, and German with the Library of Congress. https://doi.org/10.1080/19349637.2012.671053

Kreis, M. C. (November 5, 2019). *Life satisfaction survey is producing welcome results for sisters*. Global Sisters Report. A project of National Catholic Reporter. Retrieved from https://www.globalsistersreport.org/news/religious-life/column/life-satisfaction-survey-producing-welcome -results-sisters

Kreis, M. C. (2020). Younger sisters' hesitations, obstacles, and fears. *Doctrine & Life, 70*(2), 50-60.

Kreis, M. C., & Crammond, J. (2019). Younger sisters' concerns prior to their entrance to religious life. *Journal of Spirituality in Mental Health, 23*(1), 47-61. https://doi.org/10.1080/19349637.2019.1626319

Kreis, M. C., Crammond, J., & Reynolds, C. (December 6, 2018). *Life Satisfaction Scale for Apostolic Women Religious (LSSAWR)*. Examiner's Manual (English).

Kreis, M. C., Crammond, J., & Reynolds, C. (January 4, 2019). *Life Satisfaction Scale for Apostolic Women Religious (LSSAWR)*. Examiner's Manual (Spanish translation).

Kreis, M. C., Crammond, J., & Reynolds, C. (May 28, 2019). *Life Satisfaction Scale for Apostolic Women Religious (LSSAWR)*. Examiner's Manual (German translation).

Kreis, M. C., & Diaz, M. K. (2021). Factors cultivating well-being of women religious in ministry and their resonance with research on the workforce. *Journal of Spirituality in Mental Health*. https://doi.org/10.1080/19349637.2021.1886622

Life Satisfaction Scale for Apostolic Women Religious (LSSAWR) website. Retrieved from https://ctu.edu/thriving-in-ministry

Lin, X., Chen, H., Tan, Y., Yang, X., & Chi, P. (2021). The empathy questionnaire for children and adolescents (EmQue-CA) in Chinese: Psychometric properties and measurement invariance. *Asian Journal of Psychiatry, 63*. https://doi.org/10.1016/j.ajp.2021.102775

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543. https://doi.org/10.1007/BF02294825

Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus User's Guide* (7th ed.). Los Angeles: Muthén & Muthén.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Razmus, M., Razmus, W., Tylka, W. T., Jović, M., Jović, M., & Namatame, H. (2020). Cross-cultural measurement invariance of the Body Appreciation Scale-2 across five countries. *Body Image, 34*, 270-276. https://doi.org/10.1016/j.bodyim.2020.07.003

Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement, 74*, 31-57. https://doi.org/10.1177/0013164413498257

Rutkowski, L., & Svetina, D. (2017). Measurement invariance in international surveys: Categorical indicators & fit measure performance. *Applied Measurement in Education, 30*, 39-51. https://doi.org/10.1080/08957347.2016.1243540

Schmitt, N., Golubovich, J., & Leong, F. T. (2011). Impact of measurement invariance on construct correlations, mean differences, and relations with external correlates: an illustrative example using Big Five and RIASEC measures. *Assessment, 18*, 412-427. https://doi.org/10.1177/1073191110373223

Schumm, W. R., Nichols, C. W., Sheetman, K. L., & Grigsby, C. C. (1983). Characteristics of responses to the Kansas Marital Satisfaction Scale by a sample of 84 married mothers. *Psychological Report, 53*, 567-572. https://doi.org/10.2466/pr0.1983.53.2.567

Schweig, J. (2014). Cross-Level Measurement Invariance in School and Classroom Environment Surveys: Implications for Policy and Practice. *Educational Evaluation and Policy Analysis, 36*(3), 259-280. https://doi.org/10.3102/0162373713509880

Sischka, P. E., Costa, A. P., Steffgen, G., & Schmidt, A. F. (2020). The WHO-5 well-being index – validation based on item response theory and the analysis of measurement invariance across 35 countries. *Journal of Affective Disorders Reports, 1*. https://doi.org/10.1016/j.jadr.2020.100020

Sorrel, M. A, García, L. F., Aluja, A., Rolland, J. P., Rossier, J., Roskam, I., & Abad, F. J. (2021). Cross-cultural measurement invariance in the Personality Inventory for DSM-5. *Psychiatry Research, 304*. https://doi.org/10.1016/j.psychres.2021.114134

Steinmetz, H. (2013). Analyzing observed composite differences across multiple groups: Is partial measurement invariance enough? *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9*, 1-12. https://doi.org/10.1027/1614-2241/a000049

Strauss, W., & Howe, N. (1991). *Generations: The history of America's future, 1584 to 2069*. New York: William Morrow and Company, Inc.

Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology, 48*, 991-1008. https://doi.org/10.1177/0022022117717028

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-Group invariance with categorical outcomes using updated guidelines: An illustration using Mplus and the lavaan/sem tools packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 111-130. https://doi.org/10.1080/10705511.2019.1602776

Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research, 78*, 341-360. https://doi.org/10.1007/s11205-005-1037-5

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4-70. https://doi.org/10.1177/109442810031002

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In K. J. Bryant, M. E. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association. https://doi.org/10.1037/10222-009

Wittberg, P. (1994). *The rise and decline of Catholic religious orders: A social movement perspective*. Albany: SUNY Press.

Wu, A. D., Li, Z., & Zumbo. B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research, and Evaluation, 12*(3). Retrieved from https://scholarworks.umass.edu/pare/vol12/Iss1/3

Yap, S. C. Y., Donnellan, M. B., Schwartz, S. J., Kim, S. Y., Castillo, L. G., Zamboanga, B. L., Weisskirch, R. S., Lee, R. M., Park, I. J. K., Whitbourne, S. K., & Vazsonyi, A. T. (2014). Investigating the structure and measurement invariance of the Multigroup Ethnic Identity Measure in a multiethnic sample of college students. *Journal of Counseling Psychology, 61*(3), 437-446. https://doi.org/10.1037/a0036253