

A Stairway to Heaven? Structure of the Religious Involvement Inventory and Spiritual Well-Being Scale

Alan J. Gow · Roger Watson · Martha Whiteman · Ian J. Deary

Published online: 8 July 2010
© Springer Science+Business Media, LLC 2010

Abstract Being religious or having spiritual beliefs has been linked to improved health and well-being in several empirical studies. Potential underlying mechanisms can be suggested by psychometrically reliable and valid indices. Two self-report measures of religiosity/spirituality were completed by a cohort of older adults: the Religious Involvement Inventory and the Spiritual Well-being Scale. Both were analyzed using principal components analysis and the Mokken scaling procedure. The latter technique examines whether items can be described as having a hierarchical structure. The results across techniques were comparable and hierarchical structures were discovered in the scales. Analysis of the hierarchy in the RII items suggested the latent trait assesses the extent to which an individual's belief in God influences their life. Examining scales with a range of psychometric techniques may give a better indication of the latent construct being assessed, particularly the hierarchies within these which may be of interest to those investigating religiosity-health associations.

Keywords Principal components analysis · Mokken scaling procedure · Religiosity · Spirituality · Hierarchical structure

Introduction

There has long been an interest in the perceived health benefits of following a religious lifestyle or having spiritual beliefs. More recently, researchers have been examining, empirically, the links between aspects of religion and spirituality and outcomes in the medical and social domains (Hummer et al. 2004; Campbell et al. 2010; Hill and Pargament 2003; Nicholson et al. 2009; Levin 2009). Whereas the indices of spiritual or

A. J. Gow (✉) · M. Whiteman · I. J. Deary
Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology,
University of Edinburgh, Edinburgh EH8 9JZ, UK
e-mail: Alan.Gow@ed.ac.uk

R. Watson
School of Nursing & Midwifery, University of Sheffield, Sheffield S10 2TN, UK

religious beliefs and involvement used in this process have often been criticized, “it is noteworthy that the construct of religion has—in spite of frequently limited and simplistic measurement—been a statistically significant factor in a myriad of studies” (Marks 2005, p. 175). In reviewing studies published in the 1990s, Chatters (2000) concluded that there was evidence of a moderate relationship between increased religious involvement and health assessed using a range of disease categories or physical indices. Across such studies, however, religiosity or spirituality covers a range of conceptualizations, from public behaviors (church membership or attendance, for example) to private beliefs and attitudes (including private prayer or ratings of religiousness) (McCullough et al. 2000).

The unadjusted results from a meta-analysis of religious involvement and survival including 42 independent effect sizes suggested that religious individuals had a 29% higher odds of survival (OR = 1.29, 95% CI 1.21–1.39; McCullough et al. 2000). However, over half of the effect sizes included “were based on single-item measures of religious attendance or subjective religiousness with limited reliability” (p. 219). Religious involvement has been variously conceptualized and McCullough et al. (2000) found that the studies using public measures reported a stronger religious involvement–mortality association; this finding may indicate potential mechanisms underlying the effect whereby public religious practices (such as attending religious services and groups associated with them) are a means of accessing advantageous psychosocial and coping resources.

The aforementioned review therefore suggests that the beneficial effect of religious involvement (at least in terms of all-cause mortality) appears to be drawn from public participation rather than attitudes or beliefs. In support of this, Levin (2009) commented as follows:

it is the eschewed communal dimension of religious participation—church or synagogue affiliation, religious service attendance, group prayer, the receipt of formal and informal religious support—that is invariably implicated as most salutary among studies of physical and mental health (p. 133).

However, single-item or brief measures of religious attendance are more commonly used and therefore more often-cited in this context than more detailed indices, which assess religious beliefs and attitudes (Hill and Pargament 2003). Observing associations between crudely defined religious attendance and health outcomes may simply mean that this behavior is the manifest marker of an underlying religiosity/spirituality, and it may be the latter which is actually driving the effect (either directly or indirectly) (Nicholson et al. 2009). Only by simultaneous assessment of a number of aspects of religious beliefs and practices is it possible to begin to examine this distinction more thoroughly. The authors recommend that future research should not simply look for an association between religious involvement and mortality, but should aim to gain a mechanistic understanding of this by using reliable, multidimensional scales.

As quoted earlier, this is an oft-cited criticism in research examining the religion/spirituality-health link (Marks 2005; McCullough et al. 2000), whereby single-item or brief measures are included within a larger battery of other behavioral indices, rather than as well-defined and potentially important predictors. Commonly, the indices used assess only religious affiliation or consist of basic measures of religious attendance (Hill and Pargament 2003). While it has been noted that the replicated association of these simple measures to a range of important outcomes is impressive, their continued use may have resulted in a poorer understanding of such associations than if “psychometrically sophisticated measures that specifically apply to health-related issues” (Hill and Pargament 2003, p. 66) were applied. Studies using solely behavioral measures of religiosity or

spirituality are unable to help tease out any mechanistic explanations, which currently remain unclear (Nicholson et al. 2009). There is an interest in examining how spirituality or religiosity might have an impact on varied aspects of people's lives, yet there is a widespread use of instruments which may not be reliable or valid for the detailed assessment of such beliefs, attitudes, or behaviors. An understanding of each scale's construction and psychometric properties is necessary in addition to the continued assessment of their reliability using diverse methods.

The current study of the psychometric properties of scales assessing religiosity and spirituality grew from an interest in measuring these constructs in an aging cohort, in the hope that they might be used longitudinally to predict important health outcomes. If we are to use such measures, it is, therefore, necessary that the measures assessing religious beliefs and practices be examined psychometrically, otherwise it then becomes more problematic when trying to uncover any potential underlying mechanisms. That is, well-defined, reliable multi-factor constructs can point toward potential mechanisms, while limited or single-item scales rarely control potential confounding from related factors. In the current study, 2 measures of religiosity/spirituality were completed by an older cohort. These were chosen to cover aspects of religiosity and spirituality including involvement, belief, and well-being. Two analytical approaches were used to investigate the psychometric properties of the scales: principal components analysis (PCA) to investigate the suggested factor structure of the scales and the Mokken scaling procedure (MSP). The latter approach seeks to investigate the hierarchy that exists within a scale and may be particularly relevant to measures of religiosity and spirituality. It is applied to these measures for the first time. PCA will be familiar to readers but a description of the MSP is now provided.

Mokken Scaling

Mokken scaling (van Schur 2003) is one of two types of item response theory (IRT), the other being Rasch modeling. Unlike classical test theories (factor analysis and internal consistency) that rely solely on covariance between items, Mokken scaling is concerned with searching for hierarchies of items in multivariate databases and establishing their validity through a series of parameters, some of which are unique to Mokken scaling. Mokken scaling is, essentially, a stochastic version of Guttman scaling, on which it is based (Mokken and Lewis 1982), and the nature and utility of hierarchical scales can be exemplified by considering a commercial aircraft pilot who is licensed to fly the largest passenger jets such as the newest double-decker airbus. Since pilots progress through the different, larger and more sophisticated aircraft, it can be assumed that such a pilot is also qualified to fly jumbo jets and every other type of commercial aircraft from single seat aircraft, through light passenger planes, small jets and so on. On the other hand, a pilot who is only licensed to fly smaller passenger jets, while able to fly all smaller craft, will not be able to fly larger craft. In other words, based on qualifications, there is a hierarchy of 'pilotness'—ability to fly (which would be the latent trait) particular aircraft—and a pilot's position on that hierarchy, which could be assigned a score, also provides a description of ability above and below that point. The aforementioned is clearly based on an example of a skill-related process; however, hierarchical structure can be usefully examined within other more trait-like psychological and social domains, for example personality (Watson et al. 2007, 2008b), depression (Watson et al. 2008a), dysphonia (Deary et al. 2010) and neurotic disorder (Bedford et al. 2010).

The parameters whereby a Mokken scale is judged to be valid include the Loevinger's coefficient (H), which is a measure of the scalability of items: the extent to which they conform to a Guttman hierarchy. Violations of Guttman hierarchy lower the value of H and $H > .3$ is considered to indicate a good Mokken scale (van Schur 2003). The reliability of a Mokken scale can be evaluated using a test–retest procedure akin to Cronbach's alpha (Watson et al. 2007) which generates a value (Rho) which should exceed .7 for a reliable scale. The probability (at $P < .05$) of obtaining a Mokken scale can be estimated allowing for multiple comparisons using a Bonferroni type procedure (Molenaar and Sijtsma 2000). Finally, the extent to which the items on the scale show invariant ordering can be estimated. Invariant item ordering refers to the extent to which items do not violate monotone homogeneity and double monotonicity which are, respectively, measures of the extent to which the score on an item increases with increasing presence of the latent trait and the extent to which the item response curves—the probability distribution function for each item—do not overlap (Sijtsma and Junker 1994). Invariant item ordering is measured using a value labeler *Crit*, which is generated by the MSP and which is a combination of parameters related to Guttman violations by items using the H values of all the items in the scale. On the basis of *Crit* values >40 , individual items can be removed until no violations of invariant item ordering are observed.

The use of hierarchical scales in social, psychological, medical and nursing research is well established (Watson 1996; Kempen and Suurmeijer 1991; Kingshott et al. 1998; Ringdal et al. 2003). An additional dimension is added to a scale if hierarchical properties are established and offer an alternative, simply, to summing Likert-type responses. The demonstration of hierarchical properties indicates that, relative to one another, items are ordered; the implication is that they are ordered along the latent trait being measured. As with classical test theory, the total score from a set of hierarchically ordered items indicates the extent to which the latent trait is present or absent, the added advantage of a hierarchical scale is that any one item in the hierarchy also indicates the extent to which the latent trait is present or absent. Such a notion may be of particular interest in the field of religion or spirituality where it is possible to conceive of individuals varying considerably in the strength of their religious convictions or practices.

In addition to the aforementioned advantages of Mokken scales, they also offer the opportunity—through the selection of scalable items from large pools of items—to produce shorter scales that are more 'user friendly' in that they take less time to complete but retain or improve on the psychometric properties of the original scale. Any such advantage will increase return rates in research projects and in everyday use thus helping to establish the psychometric properties of the instrument better.

To summarize, the ambiguity or disagreement in the methodological or conceptual underpinnings of any religion-health investigation have been variously highlighted (Chatters 2000; Hill and Pargament 2003). It is necessary that the measures used to assess religious beliefs and practices be examined psychometrically; otherwise, it becomes more problematic when later trying to determine any potential underlying mechanisms. In the current study, 2 measures of religiosity/spirituality were completed by an elderly cohort. Two analytical approaches were used to investigate the psychometric properties of the scales: principal components analysis (PCA) and the Mokken scaling procedure (MSP).

Methods

Participants were from the Lothian Birth Cohort 1921 (LBC1921), a longitudinal study of cognitive aging. All were born in 1921 and were surviving participants of the Scottish Mental Survey of 1932. The recruitment and testing of this cohort has been described in detail previously (Gow et al. 2008). In summary, individuals were identified and recruited into the LBC1921 from 1999 to 2001. At this baseline, participants attended a clinical visit where they completed a battery of cognitive tests and a range of physical and medical assessments ($N = 550$, 234 men and 316 women: Deary et al. 2004). A second wave of follow-up was conducted from 2003 to 2005 ($N = 321$, 145 men and 176 women: Gow et al. 2008) and a third from 2007 to 2008 ($N = 237$, 109 men and 128 women: Gow et al. under review). Participants have completed several self-report questionnaires as part of, and between, these waves of testing. The current religiosity/spirituality questionnaires were completed during the second wave of follow-up when the participants were a mean age of 83.4 years ($SD = .5$). The LBC1921 participants were of either the Christian faith or none. In terms of church attendance, of the 489 who gave this information as part of an activity assessment at wave 1 (aged ~ 79 years old), the majority (223 participants or 45.6%) reported frequent church attendance, 82 participants (16.8%) sometimes attended, 86 (17.6%) rarely attended, and 98 (20.0%) never attended.

Procedure

Aspects of religiosity and spirituality were assessed by 2 subscales of the Religious Involvement Inventory (RII: Hilty and Morgan 1985) and the Spiritual Well-being Scale (SWBS: Paloutzian and Ellison 1982). These measures were chosen after consulting 'Measures of religiosity' (Hill and Hood 1999): a detailed collection of indices assessing various aspects of religious beliefs, attitudes and practices. Briefly, the scales selected were chosen for their: suitability for a UK cohort of Christian individuals, previous data concerning reliability and validity, and the assessment of various aspects of religious and spiritual beliefs and attitudes. These measures were included in a larger questionnaire booklet, which also included sections to assess occupational characteristics, lifetime activity participation and lifetime social support networks. The booklet was distributed in 2004 during the second wave of follow-up. It was sent to the 488 participants listed in the LBC1921 at that time; that is, those participants who had not withdrawn or were known to have died since the first wave of testing. Procedures were in place to follow-up non-responders or booklets received with omissions, multiple responses and incongruent answers.

Of the 488 participants mailed the booklet, 444 (91.0%) responded. This response included returned booklets from 384 participants (78.7% of those mailed), plus a booklet from a participant who had not attended the wave 1 clinical visit and was excluded. At the end of data collection, 323 booklets (84.1% of those returned) were complete and 61 (15.9%) remained partially completed after corrections were requested, where appropriate. Fifty-nine participants (12.1% of those mailed) refused the booklet for various reasons, and 44 participants (9.0% of those mailed) did not respond.

Three hundred and eighty-four participants (157 men and 227 women) attempted the religiosity/spirituality measures contained in the questionnaire booklet. Of these, 345 (89.8%) were fully completed returns while 39 (10.2%) remained partially complete.

Questionnaires

Religious Involvement Inventory (RII; Hilty and Morgan 1985). Fourteen items from the Personal Faith subscale (e.g., “It is important to me to spend periods of time in private religious thought and meditation”) and 19 items from the Orthodoxy subscale (e.g., “I know that I need God’s continual love and care”) of the RII were used. One Orthodoxy item had been removed because it was a repetition of a Personal Faith item. Item wording was altered to allow a 4-choice answer format to be used for all items (regularly to never, or strongly agree to strongly disagree, which were assigned numerical values of 3 to 0, respectively).

Spiritual Well-being Scale (SWBS; Paloutzian and Ellison 1982). The SWBS is a proprietary instrument and was reproduced with permission. The SWBS contains 20 items, 10 each for religious well-being (e.g., “I have a personally meaningful relationship with God”) and existential well-being (e.g., “I feel very fulfilled and satisfied with life”), answered on a 6-point scale from 1 (strongly disagree) to 6 (strongly agree).

Data Analysis

Data were entered into SPSS 15.0 for analysis using principal components analysis (PCA). A description of the PCA of each scale follows in the Results section. The database was then checked for missing values, for which cases were removed listwise. The resulting file was saved in SPSS in tab-delimited format with the spreadsheet option turned off and imported into the Mokken scaling procedure (MSP).

The MSP was run by entering data into the MSP software version 5.0 for Windows (Molenaar and Sijtsma 2000; iec proGamma, Groningen, the Netherlands) and setting the H value at .05 and increasing this through .05 increments to a point where no items formed scales. Between this range, all items will scale at the lower H and at approximately $H = .3$ –.4 scales will be formed. Reliable scales ($Rho > .7$) may then be investigated and refined further.

Assessing Prospective Validity

In order to assess the prospective validity of the RII and SWB scales, health-related outcomes from the wave 3 assessment were considered (when the participants were a mean age 86.6 years, $SD = .4$). These were satisfaction with life (from the 5-item Satisfaction with Life Scale; Diener et al. 1985), depression (from the Hospital Anxiety and Depression Scale; Zigmond and Snaith 1983) and lung function (as forced expiratory volume in 1 s). These were intended to indicate overall well-being, mood, and a marker of physical fitness, respectively. Separate linear regressions were run for each outcome. Firstly, the PCA-derived RII and SWB scales were entered as independent variables. The regressions were then repeated using the Mokken-derived scales. The percentages of variance accounted for were compared across analyses. [Note, as examining the psychometric properties of the scales was the primary interest, the purpose of the regression analysis was to compare the performance of the PCA versus Mokken-derived religiosity/spirituality scales, rather than against predictors from other domains.]

Table 1 Fifteen-item RII Mokken factor

Item	Mean	<i>H</i>
Property (house, automobile, money, investments, etc.) belongs to God; we only hold it in trust for Him	2.17	.76
It is important to me to spend periods of time in private religious thought and meditation	2.28	.79
I know that God answers my prayers	2.42	.84
The church is important to me as a place where I get strength and courage for dealing with the trials and problems of life	2.49	.83
I believe in salvation as release from sin and freedom for new life with God	2.50	.85
I often ask God to forgive my sins	2.51	.83
Religion is especially important to me because it answers many questions about the meaning of life	2.51	.87
I frequently feel very close to God in prayer, during public worship, or at important moments in my daily life	2.53	.87
Religion is important in my life today	2.58	.84
God has influenced my life	2.62	.83
I believe that Christ is a living reality	2.67	.84
I believe in God as Heavenly Father who watches over me and to whom I am accountable	2.73	.86
I believe that the word of God is revealed in the Scriptures	2.77	.84
I know that I need God's continual love and care	2.77	.89
I believe that God revealed Himself to man in Jesus Christ	2.80	.84

Scale *H* = .84; reliability = .98; *P* = .00026 (*n* = 356)

Results

Religious Involvement Inventory

A PCA was conducted on the 33 Religious Involvement Inventory (RII) items. The overall MSA was .98 and the lowest individual item MSA was .95; it was, therefore, not necessary to exclude any items at this stage. The scree plot and 'Eigenvalues greater than 1' criterion suggested that only a single component, explaining 62.2% of the total variance, should be extracted. This 1st unrotated component was characterized by all 33 RII items loading over .30 (available on request). They were summed to give a religious involvement inventory score, with a very high internal consistency (Cronbach's alpha = .98). When the 33 RII items were summed to create 2 subscales as previously described (Hilty and Morgan 1985; Hummer et al. 2004)—personal faith (14 items) and orthodoxy (20 items)¹—these correlated .86 (*P* = .000). This suggests the 2 RII subscales share at least 74% of their variance and so may not be validly separable. Cronbach's alpha coefficients were .94 for personal faith and .98 for orthodoxy, much higher than the previously published values of .87 and .85, respectively (Hill and Hood 1999).

The results of the MSP are shown in Table 1. The Mokken analysis of the RII accords with the previous PCA, suggesting the presence of a single scale but with fewer items. The

¹ Note that this totals 34 items, however, as one item is shared by both factors the actual number of items completed is 33.

Loevinger's coefficient was acceptable ($H = .84$), and the scale was reliable ($Rho = .98$) with a probability, corrected for multiple comparisons, of .00026. The scale runs from the 'least difficult' (i.e., highest mean score, most endorsed) item of "I believe that God revealed Himself to man in Jesus Christ," (mean score = 2.80) to the 'most difficult' (i.e., lowest mean score, least endorsed) item of "Property (house, automobile, money, investments, etc.) belongs to God; we only hold it in trust for Him," (mean score = 2.17) through other items such as "God has influenced my life," and "I know that God answers my prayers". The hierarchy of items in the scale suggests that the latent trait being measured here assesses, in individuals, the extent to which their belief in God has an influence on their life. At the most basic level, people more readily endorse a profound belief in God (and Christ, in this case) but less readily endorse His influence on their lives and a total surrender of material goods to Him.

Spiritual Well-Being Scale

The 20 Spiritual Well-being Scale (SWBS) items were first investigated using PCA. The overall MSA was .92 with the lowest individual value being .77. The scree plot and 'Eigenvalues greater than 1' criterion suggested the extraction of 3 components (direct oblimin rotation), explaining 64.1% of the total variance (Table 2). The 1st rotated component is described by 11 items loading over .30; the 10 religious well-being items load onto this factor, plus one existential well-being item, although with a lower loading (this latter item has a comparable loading onto the 3rd rotated component). The 2nd and 3rd

Table 2 Spiritual Well-being Scale component loadings

SWBS item	Component		
	1	2	3
17	.896	-.107	.178
11	.850	-.128	.121
15	.846	-.199	.151
19	.841	-.196	.177
13	-.836	-.177	.171
7	.833	-.138	.187
3	.807	-.185	.128
9	-.804	-.176	.232
5	-.778	-.116	.197
1	-.774	-.215	.236
2	-.478	-.081	.404
10	.006	-.775	-.145
14	.017	-.738	-.207
8	-.014	-.728	-.180
4	-.011	-.683	.033
20	.338	-.551	-.017
16	.026	-.007	.719
18	-.095	.171	.696
6	.082	.178	.680
12	.032	.181	.657

Note Loadings over .3 are highlighted in bold. SWBS item numbers refer to the order they appear in the scale. Odd numbered items assess religious well-being and even numbered items assess existential well-being (Paloutzian and Ellison 1982)

Table 3 Spiritual Well-being Scale Mokken factors

	Item	Mean	<i>H</i>
Factor 1	I have a personally meaningful relationship with God	3.39	.78
	I feel most fulfilled when I'm in close communion with God	3.43	.80
	I believe that God is concerned about my problems	3.64	.79
	My relationship with God helps me not to feel lonely	3.64	.80
	My relation with God contributes to my sense of well-being	3.64	.78
	I believe that God loves me and cares about me	3.93	.77
Factor 2	I feel a sense of well-being about the direction my life is headed in	4.17	.48
	I feel good about my future	4.22	.46
	I feel very fulfilled and satisfied with life	4.24	.45
	Life doesn't have much meaning	4.53	.37
	I feel that life is a positive experience	4.59	.35
	I don't enjoy much about life	4.81	.37

Factor 1: $H = .79$; reliability = .95; $P = .00038$ ($n = 352$), Factor 2: $H = .41$; reliability = .80; $P = .0014$ ($n = 352$). The SWBS items are reproduced with permission

rotated components are described by 5 items each, and result from a split in the existential well-being items according to whether they are positively worded (e.g., “I feel a sense of well-being about the direction my life is headed in”), or negatively worded (e.g., “I feel that life is full of conflict and unhappiness”). A total SWBS score was created by summing all 20 items and had a high internal consistency (Cronbach's alpha = .91). Existential well-being (EWB) and religious well-being (RWB) scores were produced by summing the appropriate 10 items describing these scales (Paloutzian and Ellison 1982), with Cronbach's alpha of .80 and .95, respectively. The correlation between RWB and EWB was .36 ($P < .001$). Therefore, both subscales have high internal consistency and, owing to their modest correlation, appear to be validly separable.

The MSP results for the SWBS are shown in Table 3. The 1st SWBS factor (Table 3) goes from a general sense of being loved by God (mean score = 3.93) to being intimate with him (mean score = 3.39) (religious well-being items). This scale has an acceptable Loevinger's coefficient ($H = .79$) and is reliable (Rho = .95) at a probability corrected for multiple comparisons of .00038. The second factor would seem to go from general enjoyment of life (mean score = 4.81) to being fulfilled and optimistic (mean score = 4.17) (existential well-being items). This scale has an acceptable Loevinger's coefficient ($H = .41$) and is reliable (Rho = .80) at a probability corrected for multiple comparisons of .0014. Again, the Mokken analysis replicates the PCA, but extends this by uncovering hierarchical properties within each scale.

Comparison of the PCA and Mokken-Derived Scales

To provide some interim validity information on the PCA versus Mokken-derived RII scale (all 33 items versus 15, respectively), the associations between these and the religious and existential well-being factors are reported in Table 4. The association between the Mokken-derived RII and the RWB is almost identical to the value for the PCA-derived RII (.89 versus .88, both $P = .000$). The association between the Mokken-derived RII and EWB was .27, similar to the RII derived from the PCA, which was .33 ($P = .000$ for both).

Table 4 Correlation coefficients of PCA and Mokken-derived Religious Involvement Inventory and Spiritual Well-being scales

	RII PCA	RII Mokken
RWB PCA	.88	.89
RWB Mokken	.88	.89
EWB PCA	.33	.27
EWB Mokken	.25	.21

RII Religious Involvement Inventory, *RWB* religious well-being, *EWB* existential well-being, *PCA* factor score as derived from principal components analysis, *Mokken* factor score as derived from the Mokken scaling procedure. All significant at $P = .000$

The association between the Mokken-derived RWB and EWB factors was .27 ($P = .000$), slightly lower than the correlation reported earlier between the full scales ($r = .36$, $P = .000$).

Finally, separate linear regression analyses were conducted with the following outcomes: satisfaction with life, depression and lung function (FEV1). For each outcome, 2 analyses were conducted: one including the PCA-derived scales and the other with those derived from the MSP. The results are summarized in Table 5. For satisfaction with life, the MSP-derived scales accounted for 11.2% of the variance versus 7.7% from the PCA scales. EWB was the only significant predictor from the PCA model, with a standardized $\beta = .27$ ($P = .000$); however, all 3 scales contributed in the MSP model: RII standardized $\beta = -.32$ ($P = .046$), EWB standardized $\beta = .29$ ($P = .000$), RWB standardized $\beta = .34$ ($P = .040$). For depression, the MSP scales accounted for 5.5% of the variance versus 4.9% from the PCA scales. Existential well-being was the only significant predictor in

Table 5 Summary of linear regression with PCA and Mokken-derived Religious Involvement Inventory and Spiritual Well-being scales as predictors

	Satisfaction with life	Depression	Lung function
PCA-derived			
Adjusted R^2	.077	.049	.069
Standardized β			
RII	-.148 ($P = .353$)	-.123 ($P = .494$)	-.420 ($P = .019$)
EWB	.274 ($P = .000$)	-.260 ($P = .002$)	-.021 ($P = .803$)
RWB	.170 ($P = .302$)	.121 ($P = .516$)	.157 (.398)
MSP-derived			
Adjusted R^2	.112	.055	.094
Standardized β			
RII	-.323 ($P = .046$)	-.001 ($P = .995$)	-.474 ($P = .010$)
EWB	.293 ($P = .000$)	-.275 ($P = .001$)	.074 ($P = .343$)
RWB	.341 ($P = .040$)	.029 ($P = .881$)	.158 ($P = .400$)

RII Religious Involvement Inventory, *RWB* religious well-being, *EWB* existential well-being, *PCA-derived* scales as derived from principal components analysis, *Mokken-derived* scales as derived from the Mokken scaling procedure; Satisfaction with life from the Satisfaction with Life Scale (Diener et al. 1985); Depression from the Hospital Anxiety and Depression Scale (Zigmond and Snaith 1983); and Lung function as forced expiratory volume in 1 s

either model (standardized $\beta = -.28$ ($P = .001$) in the MSP model and $-.26$ ($P = .002$) in the PCA model). Finally, the MSP scales accounted for 9.4% of the variance in lung function, compared with 6.9% from the PCA scales. In both cases, RII was the only significant predictor, with standardized $\beta = -.47$ ($P = .010$) in the MSP model and $-.42$ ($P = .019$) in the PCA model.

Discussion

For both the Religious Involvement Inventory and the Spiritual Well-being Scale, the PCA and MSP analyses produced helpful and provide complementary results. The analyses suggested as follows: the 2 subscales of the RII were not distinct and that a single factor more effectively described the items; the SWBS did appear to be described by 2 factors, perhaps due to the fact the items are more distinct and clearly either detail satisfaction with religious life or the strength of beliefs in this domain versus more general feelings of life satisfaction. For both the RII and SWBS, hierarchical structures were described.

The RII did not produce the 2 expected subscales in this sample. This may be partly due to the slightly altered item format used to that suggested by the authors of the scale. That is, there is a literature that suggests alterations in item and answer format can have a major impact on the way in which participants read, understand and respond to self-report items (Schwarz 1999). We would suggest this is unlikely to be a major reason in this instance as the changes made were minimal and did not affect item content. For example, the original item “To what extent has God influenced your life?” was replaced with “God has influenced my life”; it can be seen that the latter wording is more appropriately aligned with the suggested item responses of strongly agree to strongly disagree. Full details of the alterations are available on request but note the response formats were never altered. In fact, the alterations to the items made the item and response match more closely than that which was suggested by the scale description (Hill and Hood 1999). Furthermore, it is possible that the original response format is in fact masking the presence of a single factor of religious involvement. Another possibility is that in our sample, drawn as they are from a year of birth cohort in a single geographical region, personal faith and orthodoxy may be less distinct than for people from other areas and practicing other forms of Christianity. Using the revised answer formats in other and more diverse groups would be necessary to address this possibility.

The Mokken scaling procedure was applied to these scales of religiosity and spirituality in an attempt to describe their psychometric properties further. This analytical technique is used in this domain for the first time, but complements the more traditional factor analytic techniques well. Both the RII and SWBS produce scales that can be described as having hierarchical properties. This may seem intuitive, insofar as individuals within a particular religious group will vary in the strength and level of their belief and practices. But what does the Mokken approach add to the psychometric evaluation of these scales? As with other recent studies (Watson et al. 2007, 2008a, b), the Mokken scaling approach, which is based on IRT, has added value to the classical approach by identifying sub-groups of items in larger item banks whereby latent trait scores may be related to scores on individual items. In addition, the Mokken scaling procedure finds items that have a rigor in their relationships that other types of psychometric analyses do not. In the case of the present study, the complex and highly personal phenomena of spirituality and religiosity have been investigated and new information about these has been provided. As with some other psychological constructs, measuring these is not merely a matter of summing scores on a

series of items which relate, unpredictably, to the latent trait. In the cases of the inventories analyzed here, there is a discernible pattern—a hierarchy, or indeed stairway—of items related to the latent traits. In the RII scale, which measures personal faith and orthodoxy, the hierarchy of items runs from a ‘bottom line’ of belief that must be fundamental to Christians, i.e., a belief that God is Christ and Christ is God. Such a belief is unconditional to committed Christians; “to know God is, according to many traditions, the central function of religion” (Hill and Pargament 2003, p. 67). It has been suggested that closeness to God may be one of the mechanisms through which religion/spirituality might exert a positive influence on health and well-being, perhaps via physiological mechanisms from stress reduction, reduced loneliness, increased confidence, etc. (Hill and Pargament 2003). However, from this fundamental premise, and running up the items of the scale in terms of difficulty, the items become more conditional: e.g. the truth of scripture; God answers personal prayer; God influences one’s life. It would be necessary to believe that Christ was God to believe, as a Christian, that these things held true but not necessary to believe or have experienced God’s intervention in one’s life to believe that Christ was God. Finally, in the RII hierarchy, there is a statement that pertains to all material things—specifically personal belongings such as house and car—belong to God and that God is the sole source to material sustenance. According to the hierarchy, this is the most difficult belief to endorse and, therefore, one not adhered to by all Christians.

Similarly, in the first factor of the SWBS, the items run in terms of ease of endorsement from a general belief that one is loved by God—again, a fundamental premise of Christianity and, indeed, monotheistic religions—through items which are more conditional regarding the influence of God in one’s life such as induced well-being as a result of one’s faith through to the least easily endorsed item that one has a personally fulfilling relationship with God. Again, demonstrating the hierarchical nature of this scale, it would be necessary to believe that one was loved by God to have a meaningful relationship and unlikely that one would report having a meaningful relationship without feeling such love. Such a finding requires validation in other aged cohorts, as it may be that our particular sample of elderly Scots have not, in general, experienced high levels of insecure attachment or struggle which for some individuals produces a defining moment in their religious life (Hill and Pargament 2003). Finally, in the second factor of the SWBS, which does not concern religious belief but, rather, satisfaction with life, the items run in terms of endorsement from a general sense of enjoying life through finding meaning and fulfillment to having a clear sense of direction in life. From the hierarchical perspective provided by Mokken scaling, one is unlikely to feel fulfilled and to have direction in life if it is not, first and foremost, enjoyable.

In general terms, therefore, the utility of Mokken scaling is that, to some extent, the response to single items on these questionnaires could indicate, without recourse to other items, the extent to which the latent trait is present. The further utility of this approach is that the larger inventories may be reduced in length thereby reducing the burden on respondents, especially older people who were the participants in the present study, while maintaining their reliabilities. This is useful in terms of producing shorter scales, where required, that still maintain the psychometric rigor of the scales from which they were derived, rather than recourse to single-item indices of beliefs or practices.

In terms of validation, the current data represent only a first step. It was shown that the Mokken-derived RII factor was less strongly associated to the existential well-being factor of the SWBS suggesting that the Mokken analysis has distilled a more concentrated set of items than the traditional PCA approach, and subsequently produced a factor more distinct from the life satisfaction-type aspect of the SWBS. The Mokken-derived existential well-

being and religious well-being factors were also more distinct from one another (indicated by the lower inter-correlation) than their PCA-derived counterparts. These revised versions of the SWB subscales may be more acceptable to researchers who wish to keep religious and existential well-being separable (Hill and Pargament 2003). Furthermore, the reduced scales produced from the MSP were compared with those derived from the PCA. The regression analyses suggested the MSP scales accounted for a larger percentage of the variance in the outcomes considered: life satisfaction, depression, and lung function. The differences in the variance accounted for are unlikely to be significantly different, although it is certainly a strength that the shorter scales perform at least as well as the longer scales, and are likely to be more acceptable to participants, especially if part of a larger assessment. The outcomes were selected as important indicators of health and well-being in older people, and therefore of interest to those investigating links between religiosity/spirituality and healthy aging. Again, it may be that the MSP has allowed more distilled, distinct constructs to emerge which subsequently have greater predictive power. It is of interest, however, that the RII was negatively associated with lung function and life satisfaction, whereas existential well-being was always associated with a more positive outcome. It will be interesting to explore this distinction further (in these and other well-being and health-related outcomes); the current analysis was intended only as a comparison of the scales derived by the different analytical methodologies and was not an attempt to fully utilize these as potential predictors. Such analyses would necessarily consider other potential predictors, confounders and then explore the underlying mechanisms, which naturally follows on from the current thorough psychometric examination of the scales used.

Replication is clearly required, and the aim would be to use the PCA and Mokken-derived versions of these scales to predict future well-being and health-related outcomes. The advantage, or otherwise, of discovering hierarchical structures in these scales in the prediction of future outcomes will be possible with ongoing follow-up with the LBC1921. The current analysis was, however, driven not in an attempt to create a new scale to assess spirituality or religiosity, but to subject two commonly used questionnaires to detailed psychometric scrutiny. In the case of the SWBS, the current results agree with those published previously and bring added value in the discovery of a hierarchy within the items. This may prove useful to researchers in the future interested in looking in more detail at the construct of spiritual well-being and its association to important life outcomes. Although the subscales of the RII were not found to be validly separable, the items did form a strong hierarchy and it may be that further refinement of the scale can be driven from this.

Indeed, Hill and Pargament (2003) suggest that the link between aspects of religion and spirituality and health-related outcomes is suitably well-replicated to allow researchers to turn attention to the potential explanatory mechanisms. One approach which they highlighted as advantageous included using “more finely delineated measures of these constructs which might relate more directly to physical and mental health” (p. 64), with ‘closeness to God’ suggested as one such construct. Using alternative psychometric techniques, such as Mokken scaling, may be one method of honing existing indices to underscore the aspects of religiosity and spirituality that might be important in this context. Further prospective work with the LBC1921 is planned in this regard.

It is necessary to examine such measures in this way to confirm whether the suggested factor structure actually exists in diverse samples. This is necessary in any domain relying on self-report scales and has been called for by researchers examining the links between religion and health (Chatters 2000; McCullough et al. 2000; Marks 2005). Scales that go beyond the measurement of a construct by more than single items require this validation

and indeed benefit from it. By validating measures of religiosity and spirituality, and thus important aspects of human character, potentially related to diverse outcomes, it is then possible to look, in more detail, at the determinants and consequences of these beliefs, attitudes and practices. Furthermore, in using validated multi-item scales, the potential mechanisms underlying any subsequent relationships found between the measures and the outcomes of interest might be suggested (or at least the aspects requiring more detailed follow-up).

Acknowledgments The SWBS was used with permission from the authors and items are reproduced here with their consent. AG was funded by a Royal Society of Edinburgh/Lloyds TSB Foundation for Scotland Research Studentship. The University of Edinburgh School of Philosophy, Psychology & Language Sciences and Psychology Department funded production and distribution of the questionnaire booklet. The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology is part of the cross-council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council and Medical Research Council is gratefully acknowledged.

References

- Bedford, A., Watson, R., Lyne, J., Tibbles, J., Davies, F., & Deary, I. J. (2010). Mokken scaling and principal components analyses of the CORE-OM in a large clinical sample. *Clinical Psychology and Psychotherapy*, *17*, 51–62.
- Campbell, J. D., Yoon, D. P., & Johnstone, B. (2010). Determining relationships between physical health and spiritual experience, religious practices, and congregational support in a heterogeneous medical sample. *Journal of Religion and Health*, *49*, 3–17.
- Chatters, L. M. (2000). Religion and health: Public health research and practice. *Annual Review of Public Health*, *21*, 335–367.
- Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 130–147.
- Deary, I. J., Wilson, J. A., Carding, P. N., MacKenzie, K., & Watson, R. (2010). From dysphonia to dysphoria: Mokken scaling shows a strong, reliable hierarchy of voice symptoms in the Voice Symptom Scale questionnaire. *Journal of Psychosomatic Research*, *68*, 67–71.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*, 71–75.
- Gow, A. J., Johnson, W., Pattie, A., Brett, C. E., Roberts, B., Starr, J. M. et al. (under review). Stability and change in intelligence from age 11 to ages 70, 79 and 87: The Lothian birth cohorts of 1921 and 1936. *Psychology and Aging*.
- Gow, A. J., Pattie, A., Whiteman, M. C., Whalley, L. J., Starr, J. M., & Deary, I. J. (2008). Mental ability in childhood and cognitive aging. *Gerontology*, *54*, 177–186.
- Hill, P. C., & Hood, R. W. (1999). *Measures of religiosity*. Birmingham, Ala: Religious Education Press.
- Hill, P. C., & Pargament, K. I. (2003). Advances in the conceptualization and measurement of religion and spirituality. *American Psychologist*, *58*, 64–74.
- Hilty, D. M., & Morgan, R. L. (1985). Construct validation for the religious involvement inventory: Replication. *Journal for the Scientific Study of Religion*, *24*, 75–86.
- Hummer, R. A., Ellison, C. G., Rogers, R. G., Moulton, B. E., & Romero, R. R. (2004). Religious involvement and adult mortality in the United States: Review and perspective. *Southern Medical Journal*, *97*, 1223–1230.
- Kempen, G. I., & Suurmeijer, T. P. (1991). Factors influencing professional home care utilization among the elderly. *Social Science and Medicine*, *32*, 77–81.
- Kingshott, R., Douglas, N., & Deary, I. (1998). Mokken scaling of the Epworth sleepiness scale items in patients with the sleep apnoea/hypopnoea syndrome. *Journal of Sleep Research*, *7*, 293–294.
- Levin, J. (2009). “And let us make us a name”: Reflections on the future of the religion and health field. *Journal of Religion and Health*, *48*, 125–145.
- Marks, L. (2005). Religion and bio-psycho-social health: A review and conceptual model. *Journal of Religion and Health*, *44*, 173–186.

- McCullough, M. E., Hoyt, W. T., Larson, D. B., Koenig, H. G., & Thoresen, C. (2000). Religious involvement and mortality: A meta-analytic review. *Health Psychology, 19*, 211–222.
- Mokken, R. J., & Lewis, C. (1982). A nonparametric approach to the analysis of dichotomous item responses. *Applied Psychological Measurement, 6*, 417–430.
- Molenaar, I. W., & Sijtsma, K. (2000). *MSP5 for Windows*. Groningen: iec ProGAMMA.
- Nicholson, A., Rose, R., & Bobak, M. (2009). Association between attendance at religious services and self-reported health in 22 European countries. *Social Science and Medicine, 69*, 519–528.
- Paloutzian, R. F., & Ellison, C. W. (1982). Loneliness, spiritual well-being and the quality of life. In L. A. Peplau & D. Perlman (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 224–237). New York: Wiley.
- Ringdal, G. I., Jordhøy, M. S., & Kaasa, S. (2003). Measuring quality of palliative care: Psychometric properties of the FAMCARE Scale. *Quality of Life Research, 12*, 167–176.
- Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93–105.
- Sijtsma, K., & Junker, B. W. (1994). A survey of the theory and methods of invariant item ordering. *British Journal of Mathematical Statistics and Psychology, 49*, 79–105.
- van Schur, W. H. (2003). Mokken scale analysis: Between the Guttman scale and parametric item response theory. *Political Analysis, 11*, 139–163.
- Watson, R. (1996). The Mokken scaling procedure (MSP) applied to the measurement of feeding difficulty in elderly people with dementia. *International Journal of Nursing Studies, 33*, 385–393.
- Watson, R., Deary, I., & Austin, E. (2007). Are personality trait items reliably more or less 'difficult'? Mokken scaling of the NEO-FFI. *Personality and Individual Differences, 43*, 1460–1469.
- Watson, R., Deary, I., & Shipley, B. (2008a). A hierarchy of distress: Mokken scaling of the GHQ-30. *Psychological Medicine, 28*, 575–579.
- Watson, R., Roberts, B., Gow, A., & Deary, I. (2008b). A hierarchy of items within Eysenck's EPI. *Personality and Individual Differences, 45*, 333–335.
- Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361–370.