

# Item response theory analysis of the spiritual well-being questionnaire

Rapson Gomez \*, John W. Fisher

*School of Behavioural and Social Sciences and Humanities, University of Ballarat, P.O. Box 663, Ballarat, Victoria 3350, Australia*

\* Corresponding author. Tel.: +61 03 53 279760; fax: +61 03 53 279754.  
E mail address: [r.gomez@ballarat.edu.au](mailto:r.gomez@ballarat.edu.au) (R. Gomez).

## 1. Introduction

There is now growing evidence linking spirituality with general well-being, and physical and mental health (Miller & Thoresen, 2003; Powell, Shahabi, & Thoresen, 2003; Seeman, Dubin, & Seeman, 2003). Fisher (1998) has proposed a broad-based spiritual well-being model, comprising obliquely related domains of personal, communal, environmental and transcendental spiritual well-being. More recently, Gomez and Fisher (2003) published the Spiritual Well-Being Questionnaire (SWBQ). This measure has four scales, one for each of the domains identified by Fisher. Existing data show generally good reliability and validity for this measure (Gomez & Fisher, 2003, submitted for publication). The study reported here extends this area of research by examining the psychometric properties of the four scales of the SWBQ using item response theory, a method not used previously to evaluate this measure.

Traditionally, the terms spirituality and by extension spiritual well-being have been viewed in terms of religion or religious spiritual behaviors and experiences (Powell et al., 2003). However, spirituality and spiritual well-being are not necessarily limited to only religious spiritual behaviors and experiences (Schneiders, 1986). According to the National Interfaith Coalition on Aging (NICA, 1975) spiritual well-being is the affirmation of life in relationships with oneself (personal), others (communal), nature (environment), and God (or transcendental other). Using this definition, Fisher (1998) proposed a multidimensional model of spiritual well-being. According to this model, spiritual well-being comprises four oblique primary order factors, namely personal, communal, environmental and transcendental. In this model, the term well-being does not necessarily imply positive or better well-being always, as the different spiritual well-being domains are believed to have different effects for different people at different times (Gomez & Fisher, 2003).

In Fisher's (1998) (see also Gomez & Fisher (2003)) model, the personal domain deals with how one intra-relates with oneself with regard to meaning, purpose and values in life. The communal domain expresses in the quality and depth of inter-personal relationships, between self and others, and includes love, justice, hope, and faith in humanity. The environmental domain deals with enjoyment, care and nurture for the physical and biological world, including a sense of awe, wonder and unity with the environment. The transcendental domain deals with the relationship of self with some-thing or some-one beyond the human level, such as a cosmic force, transcendent reality, or God, and involves faith towards, adoration and worship of, the source of mystery of the universe. Fisher's multidimensional model has been supported in a number of studies (Gomez & Fisher, 2003; Fisher, 1998, 2001; Fisher, Francis, & Johnson, 2000).

Recently Gomez and Fisher (2003) published the Spiritual Well-Being Questionnaire (SWBQ). The SWBQ was developed to measure the four spiritual well-being domains proposed by Fisher (1998). It comprises 20 items, with separate scales for personal, communal, environment and transcendental spiritual well-being, with five items in each scale. In a series of four studies reported in the same paper, Gomez and Fisher (2003) provided evidence for the validity of the SWBQ. Both exploratory factor analysis and confirmatory factor analysis supported the four-factor model. A joint factor analysis of the four SWBQ domains with Eysenck's personality dimensions (Eysenck & Eysenck, 1991) showed that the spiritual well-being domains were independent of the personality dimensions, thereby supporting their factorial independence. Also, consistent with predictions from existing theory and data, the SWBQ domain scores (or total scale scores) for personal, communal, and environmental spiritual well-being correlated as expected with extravert-

sion, neuroticism, psychoticism, and happiness (as measured by the Oxford Happiness Inventory). The SWBQ scale scores also contributed additional variance over that of the personality dimensions in the prediction of happiness, thereby indicating support for their incremental validity. The results also showed that SWBQ scale scores correlated appropriately with the scores of Ellison's (1983) Spiritual Well-Being Scale. In addition, the four studies supported the reliabilities of the four subscales in terms of internal consistency, composite reliability, and variance extracted. Overall, these findings indicate good support for the validity and reliability of the SWBQ.

In a subsequent study, Gomez and Fisher (submitted for publication) examined the equivalencies of the measurement and structural models of the SWBQ in females and males, and the latent mean differences in these groups for the four SWBQ scales. Multi-group confirmatory factor analysis was used. The participants were 3101 females and 1361 males, with age ranging from 15 years to 32 years. The statistical fit results supported the invariance of the measurement model, and some aspects of the structural model. The practical fit indices results provided support for the invariance of both the measurement and structural models. The results also showed gender difference for only the communal latent factor, with females scoring higher. These findings add further support for the construct validity of the SWBQ.

Overall, it would appear that the four scales of SWBQ can be considered reliable and valid scales. However, to-date the studies that have examined the psychometric properties of the SWBQ scales have all used scores based on the traditional classical test theory (CTT; Lord & Novick, 1968). Besides the CTT, another approach for examining the psychometric properties of measures is item response theory (IRT; Brinbaum, 1968; Emberston & Reise, 2000; Rasch, 1960). IRT is a model-based measurement theory that aims to show the relationship between responses to items and the ability or trait that each item is supposed to be measuring (Brinbaum, 1968; Emberston & Reise, 2000; Rasch, 1960).

In IRT, the responses to items are used to obtain continuous scaled estimates of the underlying trait, called theta ( $\theta$ ). In most computer programs, the values have a mean of zero and a standard deviation of one. Two common item parameters produced by IRT are the item difficulty parameter (also called the threshold parameter) and the item discrimination parameter (or slope). The threshold parameter ( $\beta$ ) indicates the point on the scale of the latent trait where a person has a 0.5 probability of responding positively to the item, while the item discrimination parameter ( $\alpha$ ) is the ability of an item to discriminate people at different levels of the underlying trait below and above the threshold parameter (Steinberg & Thissen, 1995). In IRT analysis, graphs of trace lines or curves are generated for each item, showing the probability of a positive response to the items as a function of the underlying trait. For an item with dichotomous responses or only two response options (such as "yes" and "no"), the trace lines are called item characteristic curves (ICCs).

IRT models also provide information functions for each item and for all items together. These are called item information function and test information function, respectively. The information function of an item indicates the reliability of an item at different points of the underlying trait, while the test information function provides the reliability of all the items together at different trait levels. IRT also provides the standard error (SE) of the test information function. As the SE of a test information function is the inverse of the test information function, the SE and the test information function can be viewed as indicators of the precision of the test at different trait levels.

It has been argued that IRT has many advantages over CTT for evaluating the psychometric properties of measures (Emberston & Reise, 2000). Three advantages are of particular relevance

to this study. Firstly, for a trait, CTT provides a single score, which is derived from the scores of the different items comprising the scale. In IRT, trait scores are available at the item level. Secondly, CTT assumes and provides only one reliability (such as internal constancy) value and one SE value for all levels of the scores obtained in a measure. In contrast, IRT provides the reliability of each item at different levels of the underlying trait, controlling for the characteristics (e.g., difficulty) of the items in the scale. Thirdly, as CTT psychometric properties, such as reliability, item-total correlation and SE are sample dependent, the properties can vary across samples. Within a linear transformation, IRT psychometric properties are assumed to be sample independent or group invariant. As IRT provides parameters at the item level, this approach would allow for the identification of items that are functioning differently in terms of their ability to discriminate, represent and reliably measure the traits at different levels of the underlying trait. This, in turn, can facilitate the development and revision of measures. Thus it can be argued that the use of IRT will not only provide more valuable data on the psychometrics of the scales and items of the SWBQ, but also provide useful directions for their improvement. Given these advantages, the aim of the current study was to use IRT to examine the psychometric properties of the personal, communal, environmental, and transcendental scales of the SWBQ.

## 2. Method

### 2.1. Participants

The sample in this study was the same sample reported in our previous paper which aimed at examining the equivalencies of the measurement and structural models of the SWBQ in females and males, and the latent mean differences in these groups for the four SWBQ scales (Gomez & Fisher, submitted for publication). In brief, the sample comprised 4464 participants from mainly secondary schools and universities, and some participants from the general community (mainly church groups) in Australia. There were also university students from the UK and Ireland. Their ages ranged from 15 years to 32 years. The sample represented approximately 70% of individuals invited to participate in the study.

### 2.2. Spiritual well-being questionnaire (SWBQ)

The SWBQ has scales for personal, communal, environmental, and transcendental spiritual well-being. In all there are 20 items, with five items for each of the four scales. To allow for self-ratings of these items, respondents are asked to indicate how they feel the statements in the items described their personal experience over the last 6 months, using a five-point Likert scale, ranging from very low (rated 1) to very high (rated 5). As already noted, the SWBQ has good reliability and validity scores from a CTT perspective.

The personal items were “*developing a sense of identity*”, “*developing self-awareness*”, “*developing joy in life*”, “*developing inner peace*”, and “*developing meaning in life*”. The communal items were “*developing a love for other people*”, “*developing forgiveness for other people*”, “*developing trust between individuals*”, “*developing respect for others*”, and “*developing kindness towards other people*”. The environmental items were “*developing connection with nature*”, “*developing awe at*

*breathhtaking view*”, “*developing oneness with nature*”, “*developing harmony with the environment*”, and “*developing a sense of ‘magic’ in the environment*”. The transcendental items were “*developing a personal relationship with God*”, “*developing worship of the Creator*”, “*developing oneness with God*”, “*developing peace with God*” and “*developing prayer life*”.

### 2.3. Procedure

The plain language statement to potential participants indicated that the study was addressing aspects of human spiritual experience and behavior. Following consent, participants were asked to complete the SWBQ either in groups at the end of lectures (for mainly university students) or at some time during school hours (for secondary school students), or individually (for mainly participants from the general community). In all instances, the completed questionnaire was collected immediately after it was completed. A total of 4572 ratings were obtained, of which there were 4462 complete ratings. Only the questionnaires with complete ratings were included in the study.

### 2.4. Statistical procedures

Currently, there are several IRT models. Among other factors, the response format of items is a critical factor for deciding which IRT model to use. An example of a model for dichotomous responses is the two-parameter logistic model (2PLM; Brinbaum, 1968). This model provides threshold and discrimination parameter estimates for items. As already noted, the SWBQ has a polytomous responses format, with the response options graded. An IRT model appropriate for this item response format is Samejima’s (1969) graded responses model (GRM). The GRM conceptualizes an item in terms of a series of  $k - 1$  or  $m_i$  response dichotomies, where  $k$  is the number of response options. Thus if there are four response options, there will be three response dichotomies, namely the first category versus all other categories, the first and second response categories versus the third and fourth response categories, and first three response categories versus the fourth category. The trace lines reflecting these comparisons, referred to as operator characteristic curves (OCCs), represent the probability of an examinee’s raw item response falling in or above a given category threshold conditional on the trait level ( $\theta$ ).

Each OCC provides the location of the appropriate threshold parameters ( $\beta$ ), which is the trait ( $\theta$ ) level where there is a 0.5 probability of endorsing the relevant response option or higher response options. The number of  $\beta$ ’s for an item will correspond to the number of response dichotomies. In the GRM, the discrimination parameter ( $\alpha$ ) for all response options of an item is constrained to be equal. This constraint is not imposed across items. Thus each item will have its own single discrimination parameter. Once the threshold and discrimination parameters for the different response dichotomies of an item are known, the probability of response to each response option in the item as a function of the underlying trait can be generated. The resulting trace lines are called category response curves (CRCs). The CRC for the first response option will be a monotonically decreasing logistic function, while the CRC for the last response option will be a monotonically increasing logistic function. The CRCs for the other response options will all be nonmonotonic logistic functions.

This study used Samejima’s (1969) GRM. All analyses were conducted with Multilog 7.0.3 (Thissen, 1991). Prior to these analyses, the unidimensionality and fit of the items in the four

subscales were examined. Another assumption in IRT is that responses to items are independent of one another or local independence (Thissen & Steinberg, 1988). This means that other than the underlying trait there is no association between the items. Unidimensionality was tested using Cronbach's alpha, item-to-total correlations, and exploratory factor analysis (Cooke & Michie, 1997; Krueger & Finger, 2001). The local independence of the items can be ascertained by factor analysis as locally dependent items will appear as a separated factor in factor analysis (Chen & Thissen, 1997). The fit for each item in terms of the GRM was examined both graphically and statistically (Drasgow, Levine, Tsein, Williams, & Mead, 1995).

### 3. Results

#### 3.1. Unidimensionality

The Cronbach's alpha values for personal, communal, environmental and transcendental were 0.81, 0.82, 0.86 and 0.95, respectively. The mean item-to-total correlations were 0.75, 0.75, 0.80 and 0.91 for personal, communal, environmental and transcendental, respectively. For each SWB scale, the factor analysis resulted in only one factor. Taken together, these findings support the unidimensionality of the four scales and the local independence of the items in each scale.

#### 3.2. Model-data fit

MODFIT was used to examine model-data fit. MODFIT computes fit plots and chi-square statistics for single items (singlets), pairs of items (doublets) and groups of three items (triplets), using the method developed by Drasgow et al. (1995). These authors have argued that using doublets and triplets provides more reliable estimates of model-data fit than singlets alone.

In the case of polytomous items, the fit-plot in MODFIT compares the observed category response curve that is estimated from the calibration sample to the corresponding empirically derived category response curve that is estimated from a cross-validation sample. Close correspondence between these curves suggests good model-data fit. In this study, fit-plots were computed for all items in each SWBQ scale. With five response options per item, there were a total of 25 fit-plots per scale, five for each item. Examination of the fit plots showed mostly good fit. A few plots showed small misfit, and only three plots showed substantial misfit. These were the plots for options 1, 2 and 3 of item 2 in the environmental scale. Fig. 1 shows the fit-plots for item 1 of the personal scale. These plots illustrate the plots that characterized most of the other items.

For statistical fit, MODFIT provides chi-squares, adjusted chi-squares (i.e., adjusted for sample size), and adjusted chi-square to degree of freedom ratios for singlets, doublets and triplets. Based on a series of studies, Drasgow et al. (1995) have shown that good fitting models have adjusted chi-square to degree of freedom ratios of less than 3 for singlets, doublets and triplets. Table 1 shows the number of ratios above and below 3 for the adjusted chi square to the degree of freedom ratio for singlets, doublets and triplets. The adjusted chi square values were based on a cross-validation sample of 3000. The table also shows the mean and SD values for each item combination. The results suggest that there was generally poor model-data fit, especially for the items in the transcendental scale. According to Drasgow et al. (1995), all IRT models will be misspecified

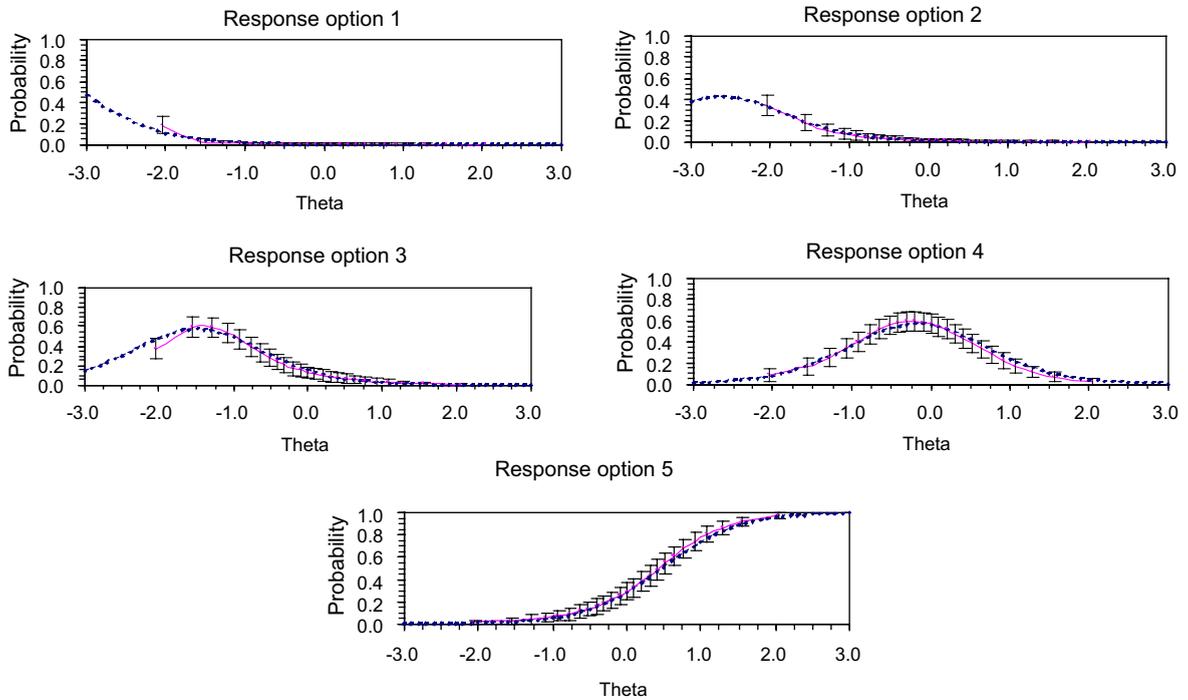


Fig. 1. Fit plots for item 1 of the personal scale (dotted lines are observed values and continuous empirical derived lines).

Table 1

Number of ratios above and below a value of 3 for the ratio of adjusted chi square to degree of freedom

	Personal				Communal			
	<3	>3	Mean	SD	<3	>3	Mean	SD
Singlets	1	4	5.11	2.68	1	4	6.02	2.77
Doublets	0	10	5.58	1.48	0	10	4.76	0.94
Triplets	1	9	4.06	0.80	3	7	3.44	0.73
	Environmental				Transcendental			
Singlets	1	4	9.17	4.55	0	5	22.50	5.76
Doublets	0	10	7.47	1.82	0	10	16.94	3.46
Triplets	0	10	4.04	0.49	0	10	10.19	1.61

to some degree, and thus test of significance for large samples (as is the case in this study) will usually lead to models being rejected. In view of this, and as the graphical results showed mostly good fit, it is appropriate to infer that there was acceptable model-data fit in this study.

### 3.3. IRT analyses of the personal scale of the SWBQ

Table 2 shows the results for the discrimination parameters ( $\alpha$ s) for all the items in the personal scale of the SWBQ. As shown in this table, the  $\alpha$  values for all items were large. With the exception

Table 2

IRT parameter estimates for the scales of the spiritual well being questionnaire

Items		$\alpha$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$
<i>Personal</i>						
1	Identity	1.97 (0.06)	-3.07 (0.12)	-2.15 (0.06)	-0.83 (0.03)	0.48 (0.03)
2	Self awareness	1.92 (0.06)	-3.16 (0.12)	-2.28 (0.07)	-0.78 (0.03)	0.64 (0.03)
3	Joy in life	2.02 (0.06)	-2.87 (0.10)	-2.09 (0.06)	-0.92 (0.03)	0.26 (0.03)
4	Inner peace	2.05 (0.06)	-2.55 (0.08)	-1.70 (0.05)	-0.38 (0.03)	0.73 (0.03)
5	Meaning in life	1.66 (0.05)	-3.01 (0.11)	-2.17 (0.07)	-0.69 (0.03)	0.64 (0.03)
<i>Communal</i>						
1	Love towards people	1.39 (0.05)	-4.12 (0.20)	-3.08 (0.12)	-1.03 (0.04)	0.75 (0.04)
2	Forgive others	1.63 (0.05)	-3.38 (0.14)	-2.42 (0.08)	-0.78 (0.04)	0.75 (0.04)
3	Trust for others	1.84 (0.06)	-3.19 (0.13)	-2.48 (0.08)	-1.07 (0.04)	0.27 (0.03)
4	Respect others	2.97 (0.09)	-2.93 (0.11)	-2.32 (0.06)	-1.18 (0.03)	0.05 (0.02)
5	Kindness to others	2.78 (0.08)	-2.94 (0.11)	-2.34 (0.06)	-1.10 (0.03)	0.17 (0.02)
<i>Environmental</i>						
1	Connect to nature	2.72 (0.07)	-1.87 (0.05)	-1.03 (0.03)	0.12 (0.02)	1.08 (0.03)
2	Awe at view	1.43 (0.09)	-2.67 (0.10)	-1.63 (0.06)	-0.14 (0.03)	1.06 (0.05)
3	Oneness with nature	3.62 (0.09)	-1.49 (0.03)	-0.75 (0.02)	0.31 (0.02)	1.21 (0.03)
4	Harmony with environment	2.90 (0.07)	-1.78 (0.04)	-0.98 (0.03)	0.16 (0.02)	1.20 (0.03)
5	See magic in environment	1.90 (0.05)	-1.47 (0.05)	-0.69 (0.03)	0.42 (0.03)	1.29 (0.04)
<i>Transcendental</i>						
1	Relation with God	4.48 (0.10)	-0.97 (0.02)	-0.38 (0.02)	0.33 (0.02)	0.93 (0.02)
2	Worship of Creator	4.48 (0.10)	-0.84 (0.02)	-0.31 (0.02)	0.35 (0.02)	0.97 (0.02)
3	Oneness with God	5.55 (0.12)	-0.83 (0.02)	-0.28 (0.02)	0.39 (0.01)	0.99 (0.02)
4	Peace with God	4.72 (0.11)	-0.98 (0.02)	-0.48 (0.02)	0.20 (0.01)	0.84 (0.02)
5	Prayer	3.91 (0.09)	-0.82 (0.02)	-0.23 (0.02)	0.52 (0.02)	1.10 (0.02)

Note:  $\alpha$  discrimination parameter;  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$  threshold parameters. The scores in the brackets are the standard error values.

of item 5 (“*meaning in life*”), the  $\alpha$  values of all the items were more or less equal (and higher than item 5). Table 2 also shows the threshold parameters ( $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$ ) for all the personal items. As will be seen in this table, the  $\beta$  parameter values made noticeable increases in the level of the latent trait at each subsequent response dichotomy. Also, the trait values for  $\beta_1$ ,  $\beta_2$  and  $\beta_3$  were somewhat evenly spaced, with all values below the mean trait level. For all items, the trait values for  $\beta_4$  were only slightly above the mean trait level.

Table 3 shows the item information function values of the five items in the personal scale. As shown, for all items, the information values from trait values  $-3.0$  to  $1.0$  were relatively high. Examination of both the CRCs and the item information functions together showed that for all items, high item information function values corresponded roughly to ratings of all five response options. Another important finding was that the CRCs for all items also showed considerable overlap in the areas occupied by the category response curves for options 1 and 2. For illustration, the CRCs for the personal item with the median  $\alpha$  value is shown in Graph 1a of Fig. 2. Graph 1b of Fig. 2 shows its item information curve. Graph 1c of Fig. 2 shows the graph for the test information functions and their corresponding standard error (*SE*) values for the

Table 3  
Information for the items in the four scales of the spiritual well being questionnaire

Items		Estimated trait						
		-3	-2	-1	0	1	2	3
<i>Personal</i>								
1	Identity	1.09	1.11	1.08	1.04	0.76	0.12	0.03
2	Self awareness	1.06	1.01	1.00	0.95	0.83	0.17	0.04
3	Joy in life	1.07	1.20	1.17	1.13	0.62	0.08	0.02
4	Inner peace	0.87	1.24	1.11	1.18	1.00	0.19	0.04
5	Meaning in life	0.78	0.80	0.78	0.78	0.65	0.18	0.05
<i>Communal</i>								
1	Love towards people	0.56	0.49	0.54	0.52	0.48	0.20	0.08
2	Forgive others	0.78	0.73	0.73	0.72	0.65	0.21	0.07
3	Trust for others	1.00	0.93	0.96	0.92	0.56	0.09	0.02
4	Respect others	2.33	2.13	2.21	2.26	0.47	0.02	0.00
5	Kindness to others	2.08	1.86	2.02	1.95	0.64	0.03	0.00
<i>Environmental</i>								
1	Connect to nature	0.31	1.87	2.07	2.00	1.99	0.32	0.04
2	Awe at view	0.51	0.62	0.59	0.61	0.57	0.28	0.11
3	Oneness with nature	0.06	1.54	3.47	2.85	3.26	0.34	0.02
4	Harmony with environment	0.22	1.94	2.34	2.19	2.19	0.41	0.04
5	See magic in environment	0.18	0.72	1.09	1.05	1.06	0.47	0.13
<i>Transcendental</i>								
1	Relation with God	0.00	0.19	5.24	4.69	5.07	0.07	0.00
2	Worship of Creator	0.00	0.11	4.54	4.97	5.20	0.08	0.00
3	Oneness with God	0.00	0.04	6.12	6.24	7.93	0.04	0.00
4	Peace with God	0.00	0.18	5.88	5.47	4.91	0.04	0.00
5	Prayer	0.00	0.15	3.44	3.34	4.23	0.21	0.01

Note:  $\alpha$  discrimination parameter;  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$  threshold parameters. The scores in the brackets are the standard error values.

personal SWB scale. As can be seen in this graph, the SE values for this scale were relatively low for trait values ranging from  $-3.0$  to  $1.0$ .

### 3.4. IRT analyses of the communal scale of the SWBQ

Table 2 also shows the results for the discrimination parameters for all the items in the communal scale. The  $\alpha$  values for all items were large. The  $\alpha$  values for items 4 (“respect for others”) and 5 (“kindness to others”) were much larger than the other three items. The  $\alpha$  values for items 2 (“forgive others”) and 3 (“trust for others”) were about the same and somewhat above the value for item 1 (“love towards others”). Table 2 also provides the  $\beta$  parameter values for all five communal items. The trait values for  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$  were quite evenly spaced, with  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  values well below the mean trait level. While the  $\beta_4$  values for all items were above the mean trait level, only items 1 (“love towards people”) and 2 (“forgive others”) showed a noticeable distance from the mean trait of 0.

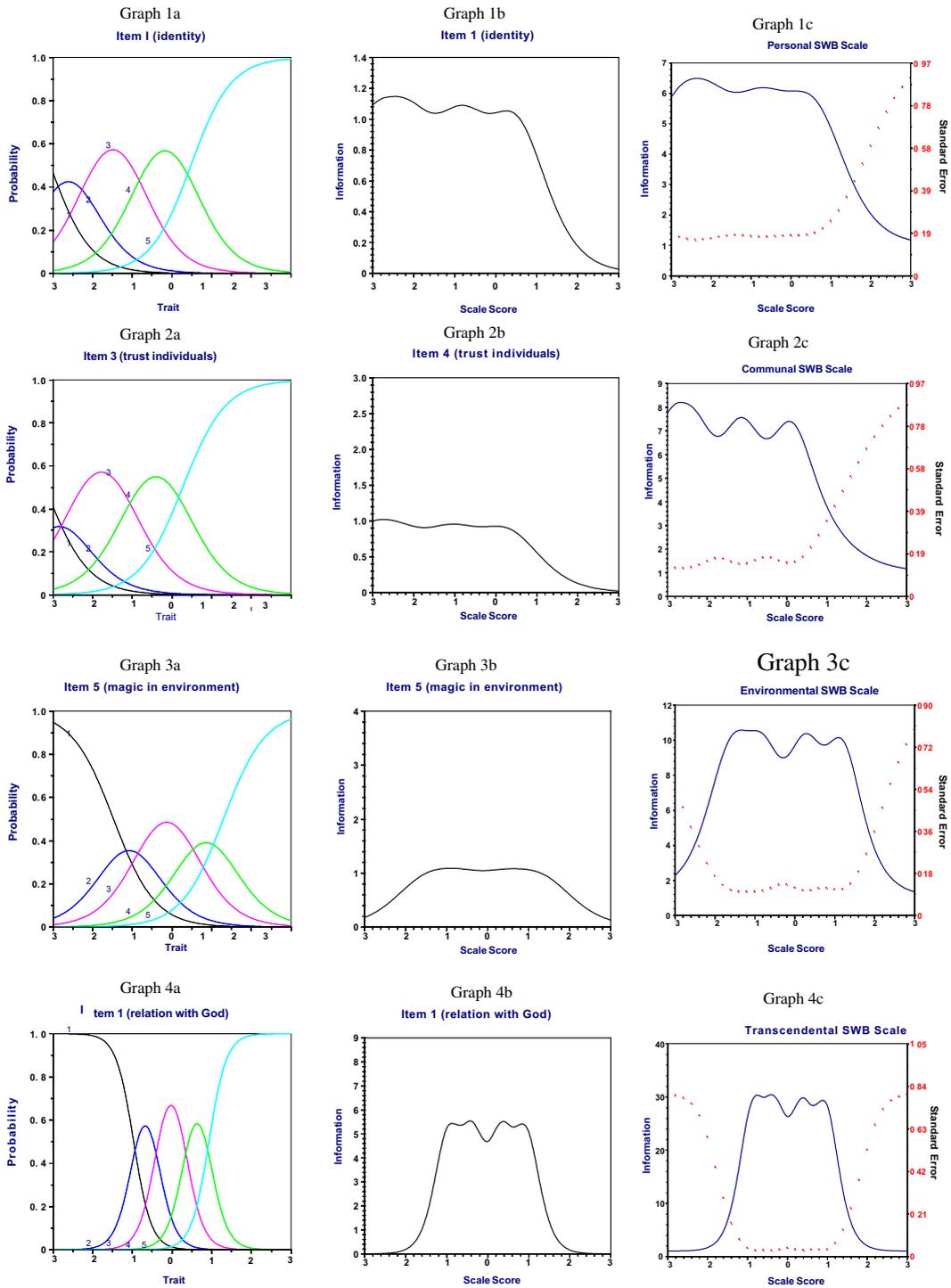


Fig. 2. Illustration of category response and item response functions, and test information function (continuous lines) and standard error curves for the four scales of the Spiritual Well Being Questionnaire.

Table 3 also includes the item information function values of the items in the communal scale. As shown, the information values of the first three items (“*love for others*”, “*forgiving others*” and “*trusting others*”) were quite low at all trait levels. The remaining two items (“*respect for others*” and “*kindness to others*”) had high information values from trait values  $-3.0$  to  $1.0$ . Examination of the CRCs and the item information functions together showed that for virtually all items, high item information function values corresponded roughly to ratings of all the five response options. Also, for all items there was considerable overlap in the areas occupied by the category response curves for options 1 and 2. For illustration, the CRCs for the communal item with the median  $\alpha$  value is shown in Graph 2a of Fig. 2, with Graph 2b showing its item information curve. Graph 2c of Fig. 2 shows the test information function and their SE values for the communal scale. As shown, the SE values for this scale were relatively low for trait values ranging from  $-3.0$  to  $0.5$ .

### 3.5. IRT analyses of the environmental scale of the SWBQ

Table 2 includes the item parameters for all the items in the environmental scale. As shown in this table, although the  $\alpha$  values for all these items were large, these values were noticeably large for items 1 (“*connected with nature*”), 3 (“*oneness with nature*”) and 4 (“*harmony with nature*”). Also there was much variability across items, except items 1 (“*connect to nature*”) and 4 (“*harmony with nature*”). As shown in Table 2, for all the environmental items, the threshold parameter values indicated noticeable moderate to large increases in the level of the latent trait at each subsequent response dichotomy. For all items, the trait values for  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$  and  $\beta_4$  were quite evenly spaced, with  $\beta_1$  and  $\beta_2$  well below the mean trait level of 0,  $\beta_3$  at around the mean trait level, and  $\beta_4$  reasonably above the mean trait.

Table 3 shows the item information function values of items in the environmental scale. As shown, the information values for items 2 (“*awe at view*”) and 5 (“*magic in environment*”) were quite low at all trait levels. The other three items had high information values from around the trait level  $-2.0$  to  $1.0$ . Examination of both the CRCs and the item information functions together showed that for all items, high item information function values corresponded roughly to ratings of all five response options. Another notable finding was that the CRCs showed considerable overlap in the areas occupied by the category response curves for options 1 and 2 for items 2 and 5. For illustration, the CRCs for the environmental item with the median  $\alpha$  value is shown in Graph 3a of Fig. 2, with Graph 3b showing its item information curve. Graph 3c in Fig. 2 shows the test information function and their corresponding SE values for the environmental scale. It shows that the SE values for the environmental scale was high for trait values ranging from  $-2.0$  to  $1.5$ .

### 3.6. IRT analyses of the transcendental scale of the SWBQ

Table 2 also includes the item parameters for all the items in the transcendental scale. As shown, the  $\alpha$  values for all items were very large. For  $\beta$  parameter values, there was noticeable moderate to large increases in the level of the latent trait at each subsequent response dichotomy (see Table 2). Also, for the all these items,  $\beta_1$  and  $\beta_2$ , values were all below the mean trait level of 0, and these values for the respective items were close to each other. The  $\beta_3$  values were only

slightly above the mean trait level. While  $\beta_4$  values were further away from the mean trait level, these values were not that different from their corresponding  $\beta_3$  values.

Table 3 includes the item information function values of items in the transcendental scale. As shown, for all items, the information values from trait values  $-1.5$  to  $1.0$  were relatively high. Examination of both the CRCs and the item information functions together showed that for all items, high item information function values corresponded roughly to ratings of all five response options. The CRCs for the item with the median  $\alpha$  value is shown in Graph 4a of Fig. 2, with Graph 4b showing its item information curve. Graph 4c shows the graph for the test information functions and their corresponding SE values for the transcendental scale. As can be seen in the graph, the SE values for this scale were relatively high for trait values ranging from  $-1.5$  to  $1.5$ .

## 4. Discussion

The study used GRM to examine the IRT psychometric properties of the items in the personal, communal, environmental and transcendental scales of the SWBQ.

### 4.1. Personal SWB scale

The results of the discrimination parameters showed that while all items in the personal scale can discriminate adequately well for high and low levels of personal spiritual well-being, the items on “*identity*”, “*self-awareness*”, “*joy in life*” and “*inner peace*” are able to do this better than the item on “*meaning in life*”. Also, the items on “*identity*”, “*self-awareness*”, “*joy in life*” and “*inner peace*” are equally good at discriminating high and low levels of personal spiritual well-being.

Overall the results for the threshold parameters showed that the items in the personal SWB scale are better able to represent the trait of personal spiritual well-being at low to the mean trait levels. Although the CRCs for all the personal items showed a corresponding noticeable shift to higher trait levels with higher response options, there was considerable overlap in the areas occupied by the category response curves for options 1 and 2 in all items. This indicates that response options 1 and 2 can be collapsed into a single response option.

In terms of reliability, the results showed that for the personal SWB scale as a whole and for the different items in this scale, there were high information values and therefore high reliability from low trait level to slightly above the mean trait level, corresponding roughly to ratings of all five response options.

### 4.2. Communal SWB scale

The discrimination parameters of the communal items showed that they can discriminate adequately well for high and low levels of communal spiritual well-being, although this ability varies across items. More specifically, the items on “*respect for others*” and “*kindness to others*” can discriminate equally well and better than the other items, and “*forgive others*” and “*trust for others*” can discriminate equally well and better than “*love towards people*”.

The findings in this study for the threshold parameters of the communal items showed that they are better able to represent the trait of communal spiritual well-being at low to the mean trait level. Also, the threshold parameters of all items in the communal SWB scale have noticeably higher trait values at each subsequent response dichotomy. However, as with the personal items, there was considerable overlap in the areas occupied by the category response curves for options 1 and 2 in all items. Thus response options 1 and 2 can be collapsed into a single response option.

The findings here also showed that the communal SWB scale as a whole and the items on “*respect for others*” and “*kindness to others*” had high information values and therefore high reliability from low trait levels to slightly above the mean trait level, corresponding roughly to ratings of all five response options. The items on “*love for others*”, “*forgiving others*” and “*trusting others*” showed quite low information values at all trait levels.

#### 4.3. Environmental SWB scale

The findings for the discrimination parameters of the environmental items suggest that although all items in the environmental SWB scale can discriminate adequately well for high and low levels of environmental spiritual well-being, their discrimination abilities vary. Also the items on “*connected with nature*”, “*oneness with nature*” and “*harmony with nature*” are better able to discriminate environmental spiritual well-being than the items on “*awe at view*” and “*see ‘magic’ in the environment*”.

The findings for the threshold parameters showed that generally the items in the environmental SWB scale are better able to represent the trait of environmental spiritual well-being at low to moderately high levels above the mean trait level. However for items 2 and 5, collapsing response options 1 and 2 into a single response option would appear appropriate. The results of the threshold parameters of all items in the environmental SWB scale showed noticeably higher values in the level of the latent trait at each subsequent response dichotomy, and their CRCs showed a corresponding shift to higher trait levels with higher response options.

For the environmental SWB scale as a whole, and for the items on “*connect with nature*”, “*oneness with nature*” and “*harmony with environment*”, there were high information values and thus high reliability from low trait levels to moderately high levels. These high values corresponded roughly to ratings of all five response options. The information values for the item on “*awe at view*” and “*magic in environment*” were quite low at all trait levels, thus these items have low reliability.

#### 4.4. Transcendental SWB scale

All items in the transcendental SWB scale had high  $\alpha$  values, suggesting that all the items in this scale can discriminate adequately well for high and low levels of transcendental spiritual well-being.

The findings for the threshold parameters of the transcendental items showed that the items in the transcendental SWB scale are better able to represent this trait from low to a point just above the mean trait level. All the items also showed noticeably higher trait values at each subsequent response dichotomy. For all items,  $\beta_1$  and  $\beta_2$  values were close to each other and well below the mean trait level. The  $\beta_3$  values were only slightly above the mean trait level, and also close the  $\beta_4$

values. Thus, it is likely that respondents rate the item on this scale in terms of three and not five response options, i.e., with both response options 1 and 2 as one option, response 3 as a second option, and both response options 4 and 5 as the third option.

For the transcendental SWB scale as a whole, and for all its items, there were high information values from moderately low trait levels to moderately high levels. The high information corresponded to responses to all the five response options.

#### *4.5. Summary and implications for revisions of the SWBQ*

Taken together, the findings can be interpreted as providing acceptability IRT based psychometric properties for all items in the personal and transcendental SWB scales. It would appear that of the five communal items, the items on “*respect for others*” and “*kindness to others*” are the only items with acceptable IRT psychometric properties. The other three items (“*love for others*”, “*forgiving others*” and “*trusting others*”) are especially weak in their reliability at all trait levels. The findings for the environmental SWB scale suggest that the items on “*connect with nature*”, “*oneness with nature*” and “*harmony with environment*” have acceptable psychometric properties from an IRT perspective. In contrast the items on “*awe at view*” and “*magic in the environment*” do not.

The study here also revealed how the SWBQ can be revised so that its psychometric properties can be improved. The communal SWB scale could benefit from a major revision. More specifically, the three items in this scale with low reliability may need to be revised to improve their reliability. For the environmental SWB scale, the two items with low reliability may need to be revised to improve their reliability. The findings here indicate that even the items identified here as having acceptable psychometric properties can benefit from some changes that will allow the items to better represent and more accurately measure the relevant traits at high levels. The findings here also indicate that the first two response options for items in the personal, communal and environmental scales could be collapsed into a single response option. Thus, it would appear that the items in these scales could be improved by having fewer response options at the lower end of the relevant spiritual well-being traits and more response options at the higher end of the relevant traits. For the transcendental items, it could appear that it may be prudent to reduce the number of response options at both the lower and higher regions of this trait. In the light of these observations, we are now in the process of revising the current edition of the SWBQ. However, this does not imply that this measure is not psychometrically sound. Indeed, the overall results of this study are supportive of its psychometric properties, as has also been the results of our two previous studies (Gomez & Fisher, 2003, submitted for publication).

In concluding, this study has shown that the use of IRT procedures can provide valuable additional psychometric information. It is well documented that good CTT based psychometric properties for a measure do not necessarily mean that it would have good IRT based psychometric properties. This has to be demonstrated using IRT procedures. This study has also demonstrated how IRT can be used to revise existing measures. It is hoped that this study has shown the value of using IRT to evaluate the psychometric properties of measures and for test development and revision, and that it will encourage other researchers to use IRT approaches for similar purposes.

## References

- Brinbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Chen, W., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289.
- Cooke, D. J., & Michie, C. (1997). An item response theory analysis of the Hare Psychopathy Checklist-Revised. *Psychological Assessment, 6*, 3–14.
- Drasgow, F., Levine, M. V., Tsein, S., Williams, B. A., & Mead, A. D. (1995). Fitting polytomous IRT models to multiple choice tests. *Applied Psychological Measurement, 19*, 143–165.
- Ellison, C. (1983). Spiritual well-being: conceptualization and measurement. *Journal of Psychology and Theology, 11*, 330–340.
- Emberston, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Eysenck, H. J., & Eysenck, S. B. G. (1991). *Eysenck personality scales (EPS adult)*. London: Hodder & Stoughton.
- Fisher, J.W. (1998). *Spiritual health: its nature, and place in the school curriculum*. Unpublished doctoral dissertation. The University of Melbourne, Melbourne, Victoria, Australia.
- Fisher, J. W. (2001). Comparing levels of spiritual well-being in state, catholic and independent school in Victoria, Australia. *Journal of Beliefs and Values, 22*, 113–119.
- Fisher, J. W., Francis, L. J., & Johnson, P. (2000). Assessing spiritual health via four domains of spiritual wellbeing: the SH4DI. *Pastoral Psychology, 49*, 133–145.
- Gomez, R., & Fisher, J. W. (2003). Domains of spiritual well-being and development and validation of the Spiritual Well-Being Questionnaire. *Personality and Individual Differences, 35*, 1975–1991.
- Gomez, R., & Fisher, J. W. (submitted for publication). The Spiritual Well-Being Questionnaire: testing for measurement and structural equivalencies and latent mean differences across gender.
- Krueger, R. F., & Finger, M. S. (2001). Using item response theory to understand comorbidity among anxiety and unipolar mood disorders. *Psychological Assessment, 13*, 140–151.
- Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Miller, W. R., & Thoresen, C. E. (2003). Spirituality, religion, and health: an emerging research field. *American Psychologist, 58*, 24–35.
- National Interfaith Coalition on Aging. (1975). *Spiritual well-being: a definition*. Athens, GA: Author.
- Powell, L. H., Shahabi, L., & Thoresen, C. E. (2003). Religion and spirituality. *American Psychologist, 58*, 36–52.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Samejima, F. (1969). Estimation of latent trait ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*. no. 17.
- Schneiders, S. (1986). Theology and spirituality: strangers, rivals or partners. *Horizons, 13*, 257–264.
- Seeman, T. E., Dubin, L. F., & Seeman, M. (2003). Religiosity/spirituality and health. *American Psychologist, 58*, 53–63.
- Steinberg, L., & Thissen, D. (1995). Item response theory in personality research. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: a festschrift honoring Donald W. Fiske*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thissen, D. (1991). *MULTILOG: multiple category item analysis and test scoring using item response theory*. Chicago: Scientific Software International, Inc..
- Thissen, D., & Steinberg, L. (1988). Data analysis using item response theory. *Psychological Bulletin, 104*, 385–395.