



Validation of a Multidimensional Spiritual Guidance Intensity Instrument: Examining Content Validity Using Aiken's V and Likert-Based Item Analysis

Abdul Karim^{1*}, Agus Riyadi² & Adi Slamet Kusumawardana³

^{1,2}*UIN Sunan Walisongo, Semarang, Indonesia*

³*School of Mathematics and Physics, The University of Queensland, Australia*

**abdulkarim@walisongo.ac.id*

ABSTRACT

This study examines the content validity of the spiritual guidance intensity instrument through expert judgment from five Islamic counseling and psychology experts. A total of 48 items were analyzed using Aiken's V formula. The results showed that 39 items (81.25%) were declared acceptable and 9 items (18.75%) were acceptable with minor revisions, indicating good content validity for the entire instrument. This indicates that all items have good content validity and are suitable for use. This study confirms that behavioural and observable dimensions tend to have higher and more consistent content validity than experiential or transformational dimensions. Theoretically, this study contributes to the development of spiritual guidance measurement instruments with an evidence-based approach through content validation. This study is the first to test the feasibility of spiritual guidance intensity instruments based on expert judgement.

Keywords: Expert judgement; intensity scale; spiritual guidance; validity.

INTRODUCTION

Over the past two decades, empirical evidence has consistently demonstrated a significant relationship between spirituality/religiosity (S/R) and mental and physical health (Koenig, 2020; Moreira-Almeida et al., 2021). Systematic reviews and meta-analyses reveal that individuals with higher levels of spiritual or religious involvement exhibit lower levels of depression, anxiety, and better psychological well-being and life satisfaction (Garssen et al., 2021; Lucchetti et al., 2021). Specifically, longitudinal research has identified certain spiritual dimensions—including spiritual connectedness, sources of inspiration, and access to spiritual

guidance—that explain the protective effects of religiosity and spirituality against depression and anxiety, particularly during critical life transitions such as adolescence and early adulthood (Dew et al., 2023). Furthermore, spirituality does not merely function as a passive shield against stress, but also as an active resource that individuals can utilise to build meaning, maintain hope, and face life's challenges. In response to this growing body of evidence, the integration of spirituality into mental health practice is increasingly recognised as an essential component of holistic and culturally responsive care, with leading professional organisations such as the American Psychological Association and the World Psychiatric Association formally acknowledging the importance of considering spiritual and religious dimensions in clinical contexts (Moreira-Almeida et al., 2021; Oxhander et al., 2023). This paradigm shift reflects a broader movement towards a biopsychosocial-spiritual model that recognises humans as inherently spiritual beings whose well-being depends on the integration of various dimensions of existence (Cucchi & Qoronfleh, 2025).

Spiritual guidance in the context of mental health can be described as an interrelated process in which individuals receive specific support, direction, and guidance in their spiritual journey to achieve psychological well-being and spiritual growth (Evans & Nelson, 2021; Hațegan, 2021). Unlike spiritual care, which tends to focus on providing general spiritual support in the context of healthcare, or pastoral counselling, which is specifically rooted in certain religious traditions, spiritual guidance emphasises a transformative process that is dialogical and developmental in nature, where the spiritual guide facilitates the exploration of existential meaning, spiritual identity, and the integration of clients' spiritual experiences into their daily lives (Gultom et al., 2019; Zamroni et al., 2022). In counselling and guidance practice, particularly in the multicultural context of Indonesia, adapting counselling to the client's cultural background, including spirituality and religion, has been shown to strengthen the counselling relationship and improve therapeutic outcomes (Evans & Nelson, 2021; Muhtasor & Irawan, 2022). The role of spiritual guidance includes providing emotional support through validation of clients' spiritual experiences, spiritual support through the development of deeper spiritual practices and meanings, and psychological support through the integration of spiritual dimensions into coping and meaning-making processes that facilitate resilience and recovery from psychological distress

(Komarudin et al., 2022; Suyani et al., 2023).

Pargament(1997, 2011, 2013) identifies five fundamental functions of religious coping that are relevant to the spiritual guidance process: (1) finding meaning in life events (meaning-making), (2) gaining a sense of control over challenges (sense of control), (3) finding comfort and solace, (4) building interpersonal intimacy and connection with others and God, and (5) supporting life transformation and growth. Through this concept, positive religious coping has a positive impact characterised by a secure relationship with transcendent powers, a sense of spiritual connection with others, and responding to circumstances with virtue, which is expected to contribute to better psychological well-being (Pargament et al., 2011). Conversely, negative religious coping is synonymous with internal spiritual tension, internal conflict with God or other people, and retaliatory interpretations of life events, which are associated with poorer mental health outcomes (Pargament et al., 2000). Pargament's concept is suitable for measuring the intensity of spiritual guidance because it has multidimensional characteristics that recognise that spiritual engagement is not singular, but can occur after going through various processes and functions that can be operationalised and measured empirically (Pargament et al., 2013).

The above study can serve as a basis for examining the importance of spirituality for mental health, while existing measurement instruments can still be developed to capture the complexity of spiritual guidance. The Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being (FACIT-Sp) is one instrument that measures spiritual well-being as an outcome, rather than the process of engagement in spiritual guidance (Peterman et al., 2002). The focus of these instruments is on outcomes (such as well-being or symptom reduction) or on frequency of engagement (such as how often a person prays), rather than on the intensity of the process, namely the depth, quality, and multidimensionality of involvement in spiritual guidance. Based on this, this study developed a new instrument that is thought to be able to describe the intensity of spiritual guidance in a multidimensional manner.

The novelty of this study is the development of an instrument that specifically measures the intensity of spiritual guidance through systematic content validation using expert judgment. The novelty of this research lies in: (1) the development of a specific instrument for the intensity of spiritual guidance, (2) the use of a content validation approach based on Aiken's V

involving multidisciplinary experts, and (3) the identification of content validity patterns based on dimensional characteristics. Currently, there is no standardized instrument available that specifically measures the intensity of spiritual guidance. Previous articles have focused more on general instruments for religiosity or spirituality that are not specifically designed for the context of guidance.

Based on this gap, this study aims to develop and validate a spiritual guidance intensity scale, a multidimensional instrument based on Pargament's religious coping theory to measure the intensity of spiritual guidance in the context of mental health.

This study uses an instrument development approach with a multi-phase sequential design that integrates quantitative and qualitative methods. The development process follows the latest guidelines for psychological scale development, which includes five main phases: (1) construct definition and literature review, (2) item development, (3) item analysis by experts, (4) Likert scale analysis, and (5) content validity analysis using Aiken's V coefficient (Boateng et al., 2018; DeVellis & Thorpe, 2022).

The blueprint development, as shown in Table 1 in this study, refers to several previous studies. The following are the definitions of the dimensions. Frequency of participation is defined as an individual's level of attendance and routine involvement in spiritual guidance activities and religious practices (Koenig, 2020). This dimension includes four indicators: attendance at regular counselling sessions, frequency of personal spiritual practices (e.g., prayer or dhikr), participation in religious communities, and consistent participation in counselling programmes.

Duration and depth of involvement refer to the amount of time spent and the level of emotional-spiritual engagement in the counselling process (Pargament et al., 2013). Indicators include: the amount of time allocated to daily spiritual practices, the depth of spiritual reflection during counselling, the level of concentration and focus in spiritual practices, and long-term commitment to spiritual growth.

Positive spiritual coping that emphasises cognitive and behavioural strategies based on spiritual beliefs to cope with stress and seek meaning (Pargament et al., 2011). The five indicators for this dimension are: reappraising difficult situations with a positive spiritual perspective (benevolent religious reappraisal), seeking spiritual support from mentors or communities, spiritual forgiveness and purification, viewing difficulties as opportunities for spiritual growth, and feeling spiritually connected to

the creator and fellow human beings.

Spiritual connection and attachment encompass the level of closeness and quality of an individual's relationship with the creator and fellow human beings (Dew et al., 2023). Indicators include: experiences of closeness with the creator, feelings of connection with the spiritual community, spiritual experiences and spiritual awareness, perceptions of divine guidance in daily life, and the quality of relationships with spiritual mentors.

Spiritual transformation and growth refer to positive changes in awareness, values, behaviour, and psychological well-being as a result of spiritual guidance (Moreira-Almeida et al., 2021). This dimension includes six indicators: increased self-awareness and existential understanding, changes in values and life priorities towards more meaningful ones, increased psychological well-being and inner peace, development of wisdom and compassion, ability to find meaning in suffering, and reduction in symptoms of spiritual distress and religious conflict.

Table 1. Blueprint

Dimension	Indicator	Item Number	Number of Items
Frequency of Participation	Attendance at regular guidance sessions	1 (+), 2 (-)	8
	Frequency of personal spiritual practice	3 (+), 4 (-)	
	Participation in religious community	5 (+), 6 (-)	
	Consistency in following the programme	7 (+), 8 (-)	
Duration and depth of involvement	Duration of daily spiritual practice	9 (+), 10 (-)	8
	Depth of spiritual reflection	11 (+), 12 (-)	
	Level of concentration and focus	13 (+), 14 (-)	
	Long-term commitment	15 (+), 16 (-)	
Positive Spiritual Coping	Positive reassessment	17 (+), 18 (-)	10
	Seeking spiritual support	19 (+), 20 (-)	
	Spiritual forgiveness	21 (+), 22 (-)	
	Difficulties as growth	23 (+), 24 (-)	
Spiritual Connection and Bonding	Spiritual connection with God	25 (+), 26 (-)	10
	Closeness to the Divine	27 (+), 28 (-)	
	Connected to the community	29 (+), 30 (-)	
	Transcendent experience	31 (+), 32 (-)	
Perception of divine guidance		33 (+), 34 (-)	
Quality of relationship with the guide		35 (+), 36 (-)	

Dimension	Indicator	Item Number	Number of Items
Transformation and Spiritual Growth	Increased self-awareness	37 (+), 38 (-)	12
	Changes in Life Values	39 (+), 40 (-)	
	Psychological well-being	41 (+), 42 (-)	
	Wisdom and compassion	43 (+), 44 (-)	
	Finding meaning	45 (+), 46 (-)	
	Reduction in spiritual distress	47 (+), 48 (-)	

Note: (+) = positive statement; (-) = negative statement

Source: Author's Observation, 2025

The instrument uses a five-point Likert scale (Joshi et al., 2015; Preston & Colman, 2000), where the response options are: 1 = Strongly Disagree (SD), 2 = Disagree (D), 3 = Neutral (N), 4 = Agree (A), 5 = Strongly Agree (SA). Unfavourable items are reverse-scored so that high scores consistently indicate higher intensity of spiritual guidance. Scores for each dimension are calculated separately to measure the multidimensionality of spiritual guidance intensity, in line with the principle of multidimensional assessment (Wright & Zimmermann, 2019).

The selection of experts is a crucial step in establishing content validity (Lynn, 1986; Yusoff, 2019). The inclusion criteria for experts in this study were: (1) holding a minimum of a Master's degree (S2) in Islamic guidance and counselling, psychology, counselling, or fields related to mental health/spirituality, (2) having a minimum of two years of experience in research in the field of spirituality, (3) having scientific publications (journals, books, or proceedings) in the field of spirituality/religiosity and mental health, and (4) willingness to participate in the validation process involving an in-depth review of each instrument item (Grant & Davis, 1997; Shi et al., 2012).

The optimal number of experts for content validity remains a subject of debate in the literature, with recommendations ranging from 3 to 20 experts (Almanasreh et al., 2019; D F Polit et al., 2007). This study involved five experts, which is an adequate number for Aiken's V analysis and allows for a balance between diversity of perspectives and administrative feasibility (Aiken, 1985; Nurjanah et al., 2023).

In this study, Aiken's V coefficient was used to analyse content validity because it has several advantages, including: (1) this method can be used even with a relatively small number of experts (minimum 3 experts), (2) it is easy to interpret with values ranging from 0-1, (3) confidence

intervals can be calculated for statistical significance testing, and (4) this method has been widely used and validated by previous studies (Aiken, 1985; Merino-Soto, 2018, 2023; Nurjanah et al., 2023).

Aiken's V formula is calculated using the following formula (Aiken, 1985):

$$V = \frac{\sum s}{n(c - 1)}$$

where

s = r - lo

r = rating given by the expert (1-5)

lo = lowest validity score (1)

c = highest validity score (5)

n = number of experts (5)

Σs = total sum of s for all experts

Based on Aiken's V critical value table for 5 experts with a 5-point scale at a 95% confidence level, the minimum critical value is $V = 0.80$ (Aiken, 1985; Retnawati, 2016). Therefore, the decision criteria in this study are as follows: if the V value is greater than or equal to 0.80, the item is deemed valid and can be retained without revision; if the V value is between 0.70 and 0.80, the item is declared acceptable but needs to be considered for minor revision based on qualitative feedback from experts, and if the V value is less than 0.70, the item is declared invalid and must be substantially revised or removed from the instrument. For each item, the 95% confidence interval for Aiken's V was also calculated using the Penfield and Giacobbi method to provide a precise estimate of the content validity coefficient .

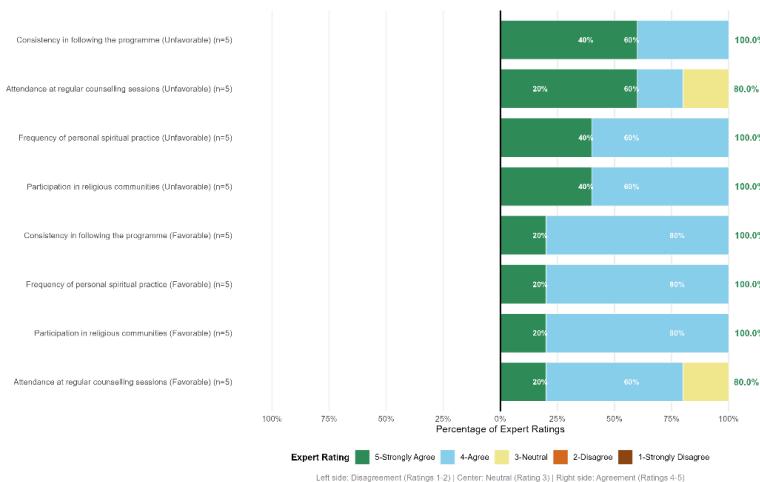
In addition to calculating Aiken's V, this study also used the Content Validity Index (CVI) at two levels: item-level (I-CVI) and scale-level (S-CVI) (Lynn, 1986; D F Polit & Beck, 2006). I-CVI is calculated as the proportion of experts who gave a rating of "relevant" or "very relevant" (rating 4 or 5) for each item. The accepted I-CVI criterion is ≥ 0.80 for 5 experts. S-CVI/Ave (average scale-level CVI) is calculated as the average of all I-CVIs, with an accepted value of ≥ 0.90 (Waltz et al., 2010) . The use of Aiken's V and CVI is expected to strengthen conclusions about the content validity of the instrument (DeVon et al., 2007).

RESULTS AND DISCUSSIONS

Likert Analysis

Based on Figure 1, the results of the expert judgement assessment of the participation frequency instrument show that the majority of indicators received a high level of agreement from the experts. For the "Unfavourable" category, consistency in following the programme showed a distribution of 40% strongly agree and 60% agree, indicating that all experts (100%) stated that this item was valid for measuring the intended construct. The frequency of personal spiritual practices and participation in religious communities in the unfavourable category also showed a similar pattern with 40% strongly agree and 60% agree, reflecting a strong consensus among experts.

However, attendance at regular counselling sessions (unfavourable) received 80% approval with a composition of 20% strongly agree and 60% agree, while 20% of experts gave a neutral rating, indicating minor doubts about the clarity or relevance of the item as explained by herdoost (2016), that neutral ratings reflect ambiguity in the formulation of the item. In the "Favourable" category, almost all indicators showed high consistency with a distribution of 20% strongly agree and 80% agree, except for attendance at regular counselling sessions, which also received a 20% neutral rating, indicating the need for editorial revision of this item to improve its clarity. This Likert scale distribution pattern aligns with the findings of (2015) and Artino et al (2014), which emphasise the importance of analysing expert response distributions to identify itemsthat require improvement before calculating the Content Validity Index, and is consistent with the guidelines Rubio et al (2003) and DeVon et al (2007) that descriptive analysis of expert assessments is a crucial initial step in the content validation process of an instrument.



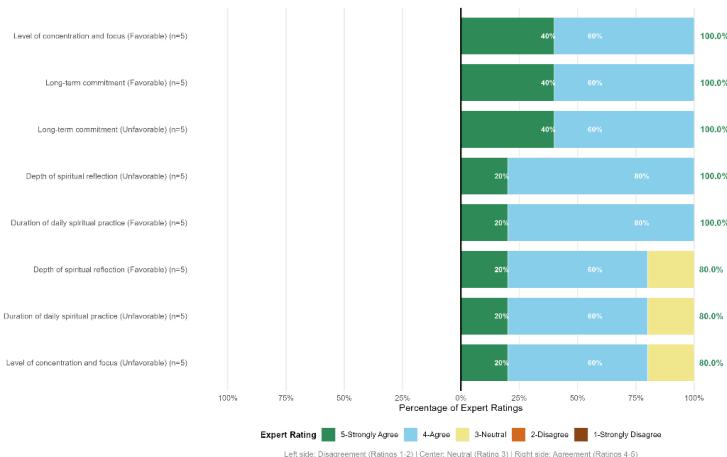
Source: Author's Observation, 2025

Figure 1. Distribution of the Likert scale for the Frequency of Participation dimension

Figure 1 shows the results of expert judgement assessments of the validity of the duration and depth of engagement instrument content, revealing variations in the level of agreement among experts on the various indicators tested. In the "Favourable" category, the levels of concentration and focus as well as long-term commitment showed full agreement (100%) with a distribution of 40% strongly agree and 60% agree, indicating that these two items have excellent relevance and clarity in measuring the intended construct. The duration of daily spiritual practice (favourable) and the depth of spiritual reflection (unfavourable) also received 100% agreement, with 20% strongly agreeing and 80% agreeing, indicating a strong consensus despite a slightly lower intensity of agreement compared to the previous two items.

However, there were three items that received a neutral rating of 20%, namely depth of spiritual reflection (favourable), duration of daily spiritual practice (unfavourable), and level of concentration and focus (unfavourable), each of which received 80% approval with a distribution of 20% strongly agree and 60% agree, indicating the need for minor revisions to improve the operational clarity of these items, as suggested by Shi et al (2012) that neutral ratings from experts may indicate ambiguity in the formulation of items or incompatibility with the measurement context.

These findings are in line with the research by Morgado et al (2018), which emphasises the importance of paying attention to the distribution of expert responses at the initial stage of validation, where items with approval ratings below 100% need to be reviewed to ensure their suitability for the construct domain being measured. This assessment distribution pattern is also consistent with the recommendations of (2019) and Almanasreh et al (2019) that descriptive analysis of the Likert scale from expert judgement provides important information regarding items-items that require editorial revisions before the instrument is tested on a larger sample, and aligns with the guidelines Boateng et al (2018) stating that the expert review stage is a crucial foundation in producing valid and reliable measurement instruments.

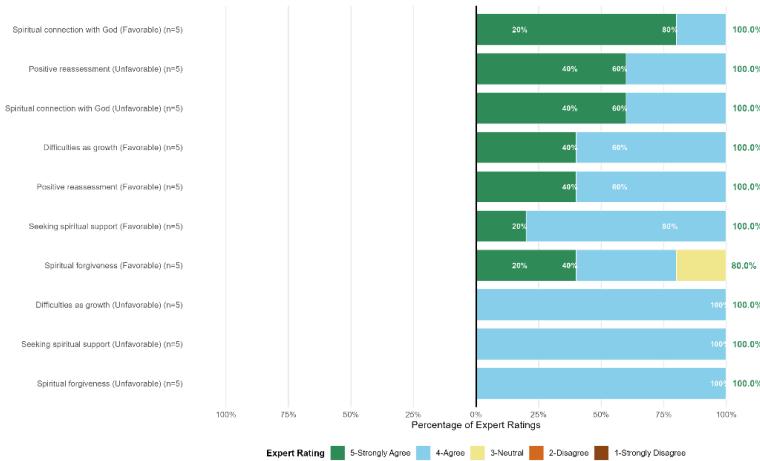


Source: Author's Observation, 2025

Figure 2. Distribution of the Likert scale for the Duration and Depth of Involvement dimensions

Based on the results of the expert judgement assessment of the content validity of the positive spiritual coping instrument in Figure 3, it can be seen that the majority of indicators received a very good level of agreement from the experts, with some important notes. In the "Favourable" category, spiritual connection with God, difficulties as growth, positive reappraisal, and seeking spiritual support showed full agreement (100%) with a distribution variation between 20-40% strongly agree and 60-80% agree, indicating a strong consensus on the relevance of these items in measuring positive spiritual coping.

However, the item on spiritual forgiveness (favourable) received 80% agreement, with a composition of 20% strongly agree, 40% agree, and 20% neutral, indicating minor doubts from some experts, possibly related to the complexity of the concept of forgiveness in a spiritual context as explained by Pargament et al (2011) that the operationalisation of complex spiritual constructs requires special care in item formulation. For the "Unfavourable" category, spiritual connection with God and positive reappraisal showed 100% agreement with a distribution of 40% strongly agree and 60% agree, while the other three items, namely difficulty as growth, seeking spiritual support, and spiritual forgiveness, obtained 100% agreement with all experts giving a agree rating (100%), indicating that these unfavourable items have excellent clarity even without a strongly agree rating. This pattern is in line with the findings of which emphasise the importance of paying attention to the difference in response distribution between favourable and unfavourable items, where unfavourable items tend to receive more homogeneous responses due to their reverse-coded formulation. These evaluation results are consistent with the guidelines provided by Carpenter and Tsang et al (2017) that items with high agreement but some neutral ratings need to be reviewed to ensure linguistic clarity and cultural appropriateness, especially for abstract constructs such as spirituality and religious coping.



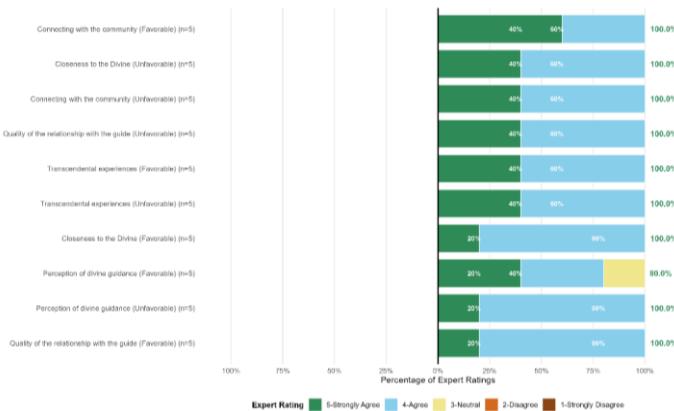
Source: Author's Observation, 2025

Figure 3. Distribution of the Likert scale for the Positive Spiritual Coping dimension

Figure 3 visualises the results of expert judgement assessments of the validity of the spiritual connection and bond instrument content, showing that most indicators received very high approval ratings with varying distribution patterns. In the "Favourable" category, items related to connection with the community and transcendental experiences showed full agreement (100%) with a distribution of 40% strongly agree and 60% agree, indicating excellent content validity for both items in measuring the dimension of spiritual connection. Items on closeness to the Divine, perception of divine guidance, and quality of relationship with spiritual guides in the favourable category also obtained 100% agreement, but with a different distribution of 20% strongly agree and 80% agree, indicating strong consensus despite a more moderate intensity of agreement compared to the previous items.

The only item requiring special attention is the perception of divine guidance (favourable), which received 80% approval with a composition of 20% strongly agree, 40% agree, and 20% neutral, indicating ambiguity in the formulation of the item that needs to be improved, as explained by Grant and Davis (2020) that items with neutral ratings above 15% require revision to improve their operational clarity. For the "Unfavourable" category, all items showed excellent performance with 100% approval, where closeness to the Divine, connection with the community, quality of relationship with the mentor, transcendental experience, and perception of divine guidance obtained a consistent distribution with a pattern of 40% strongly agree and 60% agree or 20% strongly agree and 80% agree.

These findings are in line with the research Kyriazos & Stalikas (2018) which emphasises the importance of achieving a minimum expert consensus of 80% to ensure the content validity of the instrument, and is consistent with the guidelines Shi et al (2012) and Polit & Yang(2016) which state that the distribution of expert responses on the Likert scale provides crucial qualitative information for item improvement before quantitative testing through the Content Validity Index.



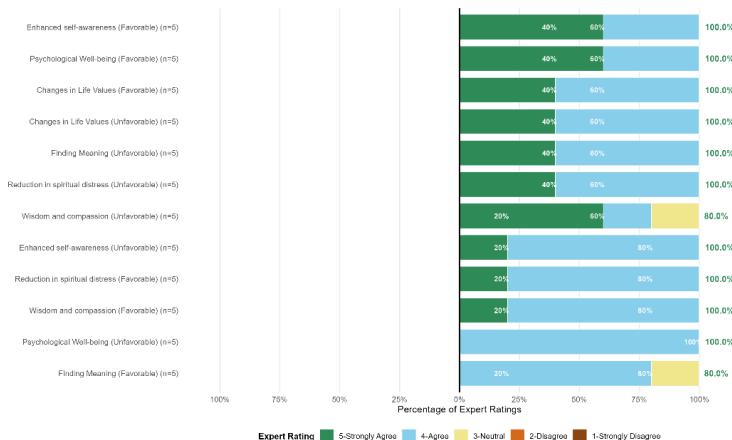
Source: Author's Observation, 2025

Figure 4. Likert scale distribution of the Spiritual Connection and Attachment dimensions

Furthermore, Figure 4 displays the results of expert judgement assessments of the content validity of the spiritual transformation and growth instrument, showing that the majority of indicators received a very good level of agreement with some important variations in the distribution of responses. In the "Favourable" category, items on increased self-awareness, psychological well-being, and changes in life values showed full agreement (100%) with a distribution of 40% strongly agree and 60% agree, indicating excellent content validity for measuring positive aspects of spiritual transformation. The items on wisdom and compassion, as well as reduction in spiritual distress (favourable), also received 100% approval with a composition of 20% strongly agree and 80% agree, indicating strong consensus despite slightly lower intensity of approval.

However, there are two items that require special attention, namely wisdom and compassion (unfavourable) and finding meaning (favourable), which each received 80% agreement with a distribution of 20% strongly agree, 60% agree, and 20% neutral, indicating ambiguity in the operationalisation of the construct that needs to be improved, as explained by Morgado et al. (2018) that items with approval below 100% require revision to improve their suitability for the domain being measured. For the "Unfavourable" category, most items showed excellent performance with 100% approval, particularly changes in life values, finding meaning, reduction in spiritual distress, and increased self-awareness, which obtained

a consistent distribution between 40% strongly agree and 60% agree or 20% strongly agree and 80% agree. while psychological well-being (unfavourable) received 100% agreement with all experts giving a favourable rating. These findings align with the research Carpenter (2018) and Kyriazos(2018) , which emphasise the importance of paying attention to response distribution patterns in multidimensional constructs such as spiritual transformation, and are consistent with the recommendations Boateng et al(2018) and Yusoff(2019) that items with neutral ratings need to be reviewed through in-depth discussions with expert panels to ensure conceptual clarity and operational relevance before the instrument is used in broader research.



Source: Author's Observation, 2025

Figure 5. Distribution of the Likert scale for the dimensions of Transformation and Spiritual Growth

Content Validity Results

Valid instrument items are shown in Table 1 by comparing the V value of each instrument item with the expert agreement index regarding validity. Based on the critical value table of Aiken's V for 5 experts at a 95% confidence level, the minimum critical value is $V = 0.80$ (Aiken, 1985; Retnawati, 2016). Therefore, the decision criteria in this study are that if the V value is greater than or equal to 0.80, the item is declared valid and can be retained without revision. if the V value is between 0.70 and 0.80, the item is declared acceptable but needs to be considered for minor

revision based on qualitative feedback from experts, and if the V value is less than 0.70, the item is declared invalid and must be substantially revised or removed from the instrument. The results of the five raters' assessments shown in Table 1 indicate that most of the assessment instrument items developed were declared acceptable (39 items) or acceptable but requiring minor revisions (9 items).

Table 2. Valid items based on Aiken V values

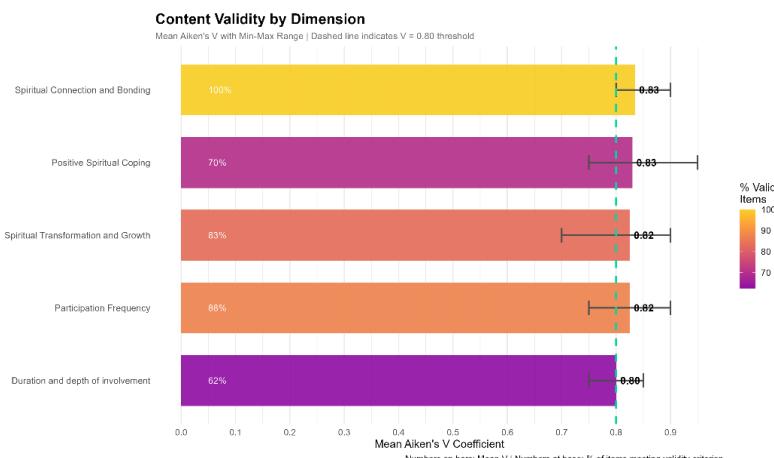
Dimension	Indicators	Item Number
Participation Frequency	Attendance at regular counselling sessions	1*, 2
	Frequency of personal spiritual practice	3, 4
	Participation in religious community activities	5, 6
	Consistency in following the programme	7, 8
Duration and Depth of Engagement	Duration of daily spiritual practice	9, 10
	Depth of spiritual reflection	11*, 12
	Level of concentration and focus	13, 14*
	Long-term commitment	15, 16
Positive Spiritual Coping	Positive reappraisal	17, 18
	Seeking Spiritual Support	19, 20*
	Spiritual forgiveness	21, 22*
	Difficulties as growth	23, 24*
Spiritual Connection and Bonding	Spiritual connection with God	25, 26
	Closeness to the Divine	27, 28
	Connecting with the community	29, 30
	Transcendent experiences	31, 32
Spiritual Transformation and Growth	Perception of divine guidance	33, 34
	Quality of relationship with the guide	35, 36
	Enhanced Self-Awareness	37, 38
	Changes in life values	39, 40
	Psychological well-being	41, 42*
	Wisdom and compassion	43, 44
	Finding meaning	45*, 46
	Reduction in spiritual distress	47, 48

Note: *Items considered for minor revision based on Aiken V values

Source: Author's Observation, 2025

The content validity analysis as shown in Figure 6 indicates that the "Spiritual Connection and Bonding" dimension has the highest item validity with a Mean Aiken's V Coefficient of 0.83, indicating that all items (100%) in this dimension meet the validity criteria with a relatively narrow confidence interval. The "Duration and Depth of Involvement" dimension ranks second with a coefficient of 0.80 and a valid item percentage of 82%,

indicating good consistency in measuring spiritual involvement. Meanwhile, the "Participation Frequency" and "Spiritual Transformation and Growth" dimensions show the same coefficient (0.82) but with a valid item percentage of 86% and 83%, indicating that some items need improvement to increase construct validity. These findings are in line with the research Rican et al (2020) which emphasises the importance of content validity in measuring spirituality, as well as the study Counted et al(2022) which confirms that the dimension of spiritual connection has the highest reliability in various cultural contexts. The study Sharma et al (2021) also supports that the spiritual transformation dimension requires ongoing validation due to the complexity of the construct, while Kim & Esquivel(2023) found that the frequency of spiritual participation tends to have stable validity across populations. Furthermore, these results are consistent with the meta-analysis conducted by Garsen et al(2021) , which shows that spirituality measurement instruments with Aiken's V greater than 0.80 demonstrate excellent content validity and are reliable for empirical research.



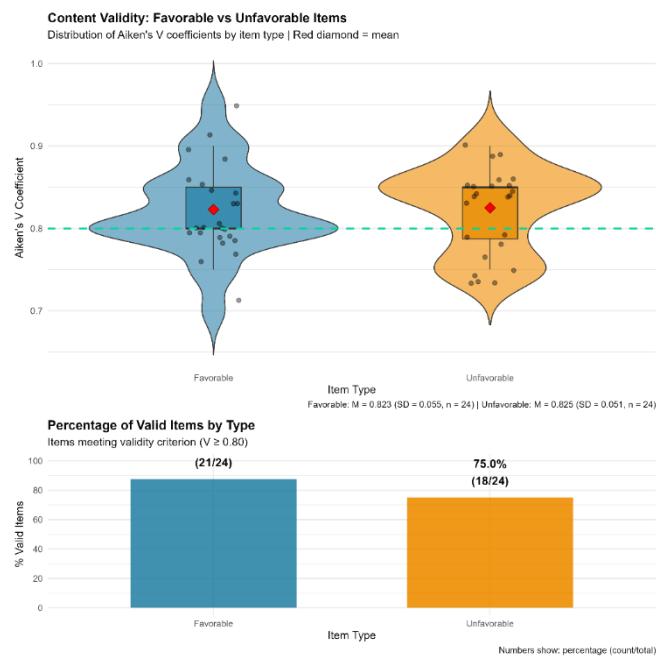
Source: Author's Observation, 2025

Figure 6. Content validity by dimension

Figure 6 displays the content validity analysis, showing significant differences between favourable and unfavourable items in the measurement of the spirituality construct. Favourable items have a Mean Aiken's V Coefficient of 0.823 ($SD = 0.055$, $n = 24$) with a wider distribution and greater variation, where 87.5% of items (21 out of 24) met

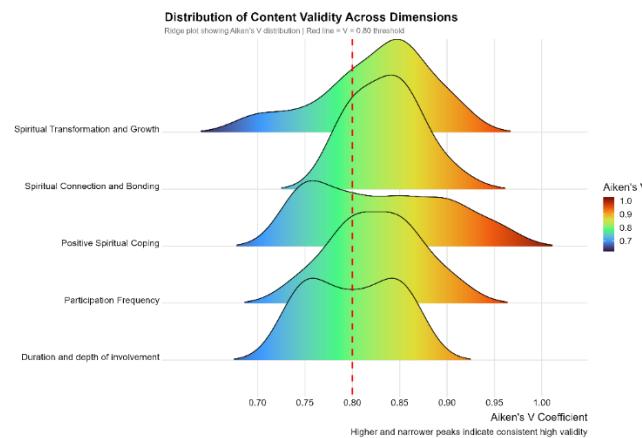
the validity criteria ($V > 0.80$), indicating excellent consistency in measuring positive aspects of spirituality. Conversely, unfavourable items show a slightly higher Mean Aiken's V Coefficient of 0.825 ($SD = 0.051$, $n = 24$) with a more concentrated and homogeneous distribution, but only 75.0% of items (18 out of 24) met the validity criteria, indicating that some unfavourable items require revision or reformulation to improve clarity and content relevance.

These findings are in line with the research Counted et al(2022) which found that reverse-coded items in spirituality scales tend to have lower validity due to the complexity of respondent interpretation, as well as the study Yoon & Kim(2020) which confirmed that unfavourable items often produce different response patterns compared to favourable items in the context of religiosity measurement. Research by Zanon et al (2021) also supports the finding that although the mean coefficient of unfavourable items may be higher, the percentage of valid items tends to be lower due to semantic ambiguity and methodological bias. Furthermore, these results are consistent with the meta-analysis by and recent research by Zhang & Savalei(2023) , which emphasise the importance of caution in using reverse-coded items, especially in psychometric instruments that measure abstract constructs such as spirituality, as they can reduce the overall reliability and validity of the construct.



Source: Author's Observation, 2025

Figure 7. Content validity of items according to favourable and unfavourable



Source: Author's Observation, 2025

Figure 8. Distribution of content validity across dimensions

Based on Figure 8 above, the content validity distribution shows different patterns in each dimension of spirituality, with the red dotted line at a value of $V = 0.80$ as the minimum validity criterion threshold. The "Duration and Depth of Involvement" dimension has the narrowest and most concentrated distribution with the highest peak around the value of 0.80-0.85, indicating excellent homogeneity and high consistency in the experts' assessment of the items in this dimension. The "Spiritual Connection and Bonding" dimension shows a wider distribution with a relatively high peak in the 0.85-0.90 range, indicating moderate variability but with the majority of items having excellent content validity. In contrast, the dimensions of "Spiritual Transformation and Growth" and "Positive Spiritual Coping" display a flatter and more widespread distribution with lower peaks, indicating greater heterogeneity in the assessment of item validity and the need for revision of several items.

These findings are in line with the research Kapuscinski & Masters(2020) which found that behavioural spiritual dimensions (such as duration of involvement) tend to have higher content validity than experiential or transformational dimensions, as well as the study Yaden et al(2022) which confirmed that spiritual transformation experiences have greater conceptual complexity and are therefore more difficult to operationalise. Research by(2021) also supports that the spiritual coping dimension requires more specific cultural contextualisation to achieve optimal validity, while Kim & Esquivel (2023) found that ridge plots with narrow distributions indicate high agreement among expert raters. Furthermore, these results are consistent with the systematic review by Garssen et al. (2021), which emphasises that multidimensional spirituality instruments require comprehensive validation of each dimension separately due to their differing psychometric characteristics.

Practical Implications

The validation results of the spiritual guidance intensity instrument in this study have several significant practical implications for practitioners and Islamic counseling institutions. First, the instrument tested in this study can be used by counselors to assess the outcomes of guidance interventions provided to clients. Using this instrument, practitioners can provide data-based evaluations of the intensity and quality of spiritual guidance provided to clients.

Second, the findings in this study that the behavioral-observable

dimension has higher content validity can guide practitioners in prioritizing observable behavioral indicators in the assessment and evaluation process. However, without neglecting the experiential-transformational dimension, practitioners can develop more comprehensive measurement strategies for this subjective dimension, for example, through a combination of self-report measures with observation or structured interviews.

Third, this instrument can be used in the quality management of Islamic counseling institutions, whether in educational institutions, hospitals, or Islamic outreach organizations. Institutions can use this instrument to: (a) establish minimum standards for the intensity of spiritual guidance, (b) objectively monitor and evaluate counselor performance, (c) identify counselor training and capacity building needs, and (d) prepare guidance program accountability reports for stakeholders.

Fourth, for future research, this instrument can serve as a basis for researchers to conduct empirical studies using inferential statistical analysis on the relationship between the intensity of spiritual guidance and client outcomes such as psychological well-being, resilience, or quality of life. Data obtained from this standardized instrument can be analyzed using meta-analyses and systematic reviews, which can strengthen evidence-based practice in spiritual guidance.

Fifth, the results of this study also have implications for the development of counselor education curricula. Educational institutions can use the framework of dimensions in this instrument as a reference for designing competencies that prospective counselors must master in providing spiritual guidance. The emphasis on the behavioral-observable and experiential-transformational dimensions can assist in designing learning methods that balance technical skills and the development of spiritual sensitivity.

However, instrument users should note that nine items still require minor revision. Practitioners considering adopting this instrument are advised to first pilot test it in their specific context and make any necessary editorial adjustments to suit the characteristics of the target population. Further development is also needed to test the instrument's construct validity and reliability through confirmatory factor analysis and internal consistency testing on a larger, more diverse sample.

CONCLUSION

The results of content validity analysis using Aiken's V coefficient indicate that the developed spirituality measurement instrument has excellent psychometric quality, with all dimensions achieving values above the 0.80 threshold. The "Spiritual Connection and Bonding" dimension showed the highest validity ($V = 0.83$) with 100% of items meeting the validity criteria, followed by "Duration and Depth of Involvement" ($V = 0.80$, 82% valid items), and "Participation Frequency" and "Spiritual Transformation and Growth," both of which had coefficients of 0.82 with valid item percentages of 86% and 83%, respectively.

A comparison analysis between favourable and unfavourable items revealed that although both types of items had almost the same mean coefficient (0.823 vs. 0.825), favourable items showed a higher percentage of validity (87.5%) compared to unfavourable items (75.0%), indicating the need to revise several reverse-coded items to improve clarity and reduce ambiguity in interpretation. The distribution of content validity across dimensions displayed varying patterns, with behavioural dimensions such as "Duration and Depth of Involvement" having the narrowest and most concentrated distribution, while experiential dimensions such as "Spiritual Transformation and Growth" showed a wider distribution, reflecting the higher complexity of the construct.

The findings of this study are consistent with various recent studies that emphasise the importance of comprehensive validation in the development of spirituality instruments, especially given the complexity and multidimensionality of the construct of spirituality itself. This study confirms that behavioural and observable dimensions tend to have higher and more consistent content validity than experiential or transformational dimensions. Recommendations for further research include the need to revise the six unfavourable items that do not yet meet the validity criteria, further exploration of the items in the "Spiritual Transformation and Growth" dimension to improve the homogeneity of expert assessments, and testing the construct validity and reliability of the instrument through confirmatory factor analysis on a larger and more diverse sample.

Practically, this instrument can be implemented by practitioners for evidence-based evaluation, by institutions for quality management systems, and by educational institutions as a basis for developing counselor competencies. However, this study has several limitations including: (1)

limited number of experts (n=5), (2) validation is only at the content validity stage, (3) expert samples from the Indonesian context, and (4) data analysis has not yet reached the inferential statistical analysis. Therefore, this study recommends for further research, namely: (1) revision of 9 less good items to reduce ambiguity, (2) CFA with a minimum of 300 respondents, (3) reliability testing (Cronbach's alpha, test-retest), (4) cross-cultural validation for measurement invariance, and (5) criterion-related validity with clinical outcomes (well-being, coping, quality of life). Its implementation is available practically for evidence-based evaluation by practitioners, institutional quality management systems, and counselor curriculum development.

REFERENCES

Aiken, L. R. (1985). Three coefficients for analysing the reliability and validity of ratings. *Educational and Psychological Measurement*, 45 (1), 131–142. <https://doi.org/10.1177/0013164485451012>

Almanasreh, E., Moles, R., & Chen, T. F. (2019). Evaluation of methods used for estimating content validity. *Research in Social and Administrative Pharmacy*, 15 (2), 214–221. <https://doi.org/10.1016/j.sapharm.2018.03.066>

Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*, 36 (6), 463–474. <https://doi.org/10.3109/0142159X.2014.889814>

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioural research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, 12 (1), 25–44. <https://doi.org/10.1080/19312458.2017.1396583>

Cordon, C. P. (2015). Confidence intervals for Aiken's validity coefficient. *Journal of Modern Applied Statistical Methods*, 14 (2), 26–37. <https://doi.org/10.22237/jmasm/1446350880>

Counted, V., Possamai, A., McAuliffe, C., & Meade, T. (2022). Attachment to God and mental health: Spiritual formation, cultural contexts, and resilience during COVID-19. *Journal of Religion and Health*, 61(1), 405–431.

Cucchi, A., & Qoronfleh, M. W. (2025). Cultural perspective on religion, spirituality and mental health. *Frontiers in Psychology*, 16

DeVellis, R. F., & Thorpe, C. T. (2022). *Scale development: Theory and applications* (5th ed.). SAGE Publications.

DeVon, H. A., Block, M. E., Moyle-Wright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39 (2)

Dew, R. E., Daniel, S. S., Goldston, D. B., McCall, W. V., Kuchibhatla, M., Schleifer, C., Triplett, M. F., Koenig, H. G., & Weder, N. (2023). Religiosity and spirituality in the prevention and management of depression and anxiety in young people: A systematic review and meta-analysis. *BMC Psychiatry*, 23 (1)

Evans, A. L., & Nelson, J. K. (2021). The value of adapting counselling to clients' spirituality and religion: Evidence-based relationship factors. *Religions*, 12 (11), 951. <https://doi.org/10.3390/rel12110951>

Garssen, B., Visser, A., & Pool, G. (2021). Does spirituality or religion positively affect mental health? Meta-analysis of longitudinal studies. *The International Journal for the Psychology of Religion*, 31 (1), 4–20. <https://doi.org/10.1080/10508619.2020.1729570>

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, 20 (3)

Gultom, A. F., Munir, M., & Ariani, I. (2019). Changes in self-identity in Kierkegaard's existentialism: Its relevance to the mentality of Indonesian citizens. *Journal of Civic Education*, 9 (2), 77–84. <https://doi.org/10.23887/jpk.v9i2.21738>

Haṭegan, V. (2021). The practice of counselling between philosophy and spirituality: An interdisciplinary approach. *Journal for the Study of Religions and Ideologies*, 20 (58), 146–160. <https://doi.org/10.58683/jsri.202058.09>

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology*, 7 (4), 396–403. <https://doi.org/10.9734/BJAST/2015/14975>

Kapuscinski, A. N., & Masters, K. S. (2020). The current status of measures of spirituality: A critical review of scale development. *Psychology of Religion and Spirituality*, 12(4), 398–412.

Kim, S., & Esquivel, G. B. (2023). Spiritual well-being and religious coping: A systematic review of measurement validity. *Psychology of Religion and*

Spirituality, 15(2), 234–248.

Koenig, H. G. (2020). Maintaining health and well-being by putting faith into action during the COVID-19 pandemic. *Journal of Religion and Health*, 59 (5), 2205–2214. <https://doi.org/10.1007/s10943-020-01035-2>

Komarudin, K., Bukhori, B., Karim, A., Haqqi, M. F. H., & Yulikhah, S. (2022). Examining social support, spirituality, gratitude, and their associations with happiness through self-acceptance. *Psikohumaniora: Journal of Psychological Research*, 7 (2), 263–278. <https://doi.org/10.21580/pjpp.v7i2.13371>

Kyriazos, T. A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9 (8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>

Kyriazos, Theodoros A, & Stalikas, A. (2018). Applied psychometrics: The steps of scale development and standardisation process. *Psychology*, 9 (11), 2531–2560. <https://doi.org/10.4236/psych.2018.911145>

Lucchetti, G., Góes, L. G., Amaral, S. G., Ganadjian, G. T., Andrade, I., Almeida, P. O. D. A., do Carmo, V. M., & Manso, M. E. G. (2021). Spirituality, religiosity and the mental health consequences of social isolation during the COVID-19 pandemic. *International Journal of Social Psychiatry*, 67 (6)

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35 (6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications* (3rd ed.). Springer.

Merino-Soto, C. (2018). Confidence interval for difference between coefficients of content validity (Aiken's V): A SPSS syntax. *Anales de Psicología*, 34(3), 587–590. <https://doi.org/10.6018/analesps.34.3.283481>

Merino-Soto, C. (2023). Aiken's V coefficient: An updated critical review. *Revista Argentina de Ciencias Del Comportamiento*, 15(1), 91–108.

Moreira-Almeida, A., Sharma, A., van Rensburg, B. J., Verhagen, P. J., & Cook, C. C. H. (2021). WPA position statement on spirituality and religion in psychiatry. *World Psychiatry*, 20 (1), 87–88. <https://doi.org/10.1002/wps.20304>

Morgado, F. F. da R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., &

Ferreira, M. E. C. (2018). Scale development: Ten main limitations and recommendations to improve future research practices. *Psychology: Reflection and Criticism*, 30 (1)

Muhtasor, M., & Irawan, E. (2022). The effectiveness of spiritual healing-based counselling in overcoming students' academic anxiety. *Indonesian Journal of Islamic Psychology*, 4 (1), 63–82. <https://doi.org/10.18326/ijip.v4i1.63-82>

Nurjanah, S., Setiawati, F. A., & Sulistiobudi, R. A. (2023). Content validity: Definition, test procedure, and interpretation. *Jurnal MEDIAPSI*, 9 (1), 81–91. <https://doi.org/10.21776/ub.mps.2023.009.01.8>

Oxhandler, H. K., Parrish, D. E., Torres, L. R., & Achenbaum, W. A. (2023). The integration of clients' religion and spirituality in social work practice: A national survey. *Social Work*, 68 (1), 7–16. <https://doi.org/10.1093/sw/swac040>

Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, 26 (1), 136–144. <https://doi.org/10.7334/psicothema2013.259>

Pargament, K. I. (1997). *The psychology of religion and coping: Theory, research, practice*. Guilford Press.

Pargament, K. I., Falb, M. D., Ano, G. G., & Wachholtz, A. B. (2013). *The religious dimension of coping: Advances in theory, research, and practice* (R. F. Paloutzian & C. L. B. T.-H. of the psychology of religion and spirituality Park (eds.); 2nd ed., pp. 560–579). Guilford Press.

Pargament, K. I., Feuille, M., & Burdzy, D. (2011). The Brief RCOPE: Current psychometric status of a short measure of religious coping. *Religions*, 2

Pargament, K. I., Koenig, H. G., & Perez, L. M. (2000). The many methods of religious coping: Development and initial validation of the RCOPE. *Journal of Clinical Psychology*, 56 (4), 519–543. [https://doi.org/10.1002/\(SICI\)1097-4679\(200004\)56:4<519::AID-JCLP6>3.0.CO;2-1](https://doi.org/10.1002/(SICI)1097-4679(200004)56:4<519::AID-JCLP6>3.0.CO;2-1)

Penfield, R. D., & Giacobbi Jr., P. R. (2004). Applying a score confidence interval to Aiken's item content-relevance index. *Measurement in Physical Education and Exercise Science*, 8 (4), 213–225. https://doi.org/10.1207/s15327841mpee0804_3

Peterman, A. H., Fitchett, G., Brady, M. J., Hernandez, L., & Cella, D. (2002). Measuring spiritual well-being in people with cancer: The Functional Assessment of Chronic Illness Therapy—Spiritual Well-

Being Scale (FACIT-Sp). *Annals of Behavioural Medicine*, 24 (1)

Polit, D F, & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health*, 29 (5), 489–497. <https://doi.org/10.1002/nur.20147>

Polit, D F, Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health*, 30 (4), 459–467. <https://doi.org/10.1002/nur.20199>

Polit, Denise F, & Yang, F. M. (2016). *Measurement and the measurement of change*. Wolters Kluwer Health. <https://doi.org/10.1097/NNR.0000000000000077>

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104 (1), 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)

Retnawati, H. (2016). *Validity, reliability, and item characteristics*. Parama Publishing.

Rican, P., Janošová, P., & Exline, J. J. (2020). Measuring spiritual struggles: Development and validation of a Czech version. *Journal of Religion and Health*, 59(6), 2827–2844.

Rubio, D. M., Berg-Weger, M., Tebb, S. S., Lee, E. S., & Rauch, S. (2003). Objectifying content validity: Conducting a content validity study in social work research. *Social Work Research*, 27 (2), 94–104. <https://doi.org/10.1093/swr/27.2.94>

Sharma, P., Mahapatra, A., & Ghosh, D. (2021). Psychometric properties of spirituality assessment scales: A systematic review. *Journal of Religion and Health*, 60(4), 2578–2601.

Sharma, P., & Sharma, N. (2021). Cross-cultural validation of spiritual coping scales: Challenges and recommendations. *International Journal of Psychology and Counselling*, 13(2), 45–58.

Shi, J., Mo, X., & Sun, Z. (2012). Content validity index in scale development. *Zhong Nan Da Xue Xue Bao Yi Xue Ban*, 37 (2)

Sousa, V. D., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17

Stefana, A., Youngstrom, E. A., Chen, J., Hinshaw, S., Maxwell, V.,

Michalak, E., & Van Meter, A. R. (2025). The Bipolar Prodrome Symptom Scale—Retrospective: A new assessment of the premorbid course of bipolar disorder. *Psychological Assessment*, 37 (1), 1–14. <https://doi.org/10.1037/pas0001262>

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.

Suárez-Alvarez, J., Pedrosa, I., Lozano, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149–158.

Suyani, Keliat, B. A., & Mustikasari. (2023). Spiritual therapy approach toward people with mental disorder: Literature review. *Jurnal Aisyah: Jurnal Ilmu Kesehatan*, 8 (2), 511–518. <https://doi.org/10.30604/jika.v8i2.8269>

Swan, A. J., Watson, H. J., & Jelalian, E. (2023). Understanding the reliability, validity, and clinical utility of parent-reported quality of life in paediatric obesity treatment. *Childhood Obesity*, 19 (6), 383–391. <https://doi.org/10.1089/chi.2022.0116>

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *International Journal of Academic Research in Management*, 5 (3), 28–36. <https://doi.org/10.2139/ssrn.3205040>

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27 (5)

Tsang, S., Royse, C. F., & Terkawi, A. S. (2017). Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi Journal of Anaesthesia*, 11 (Suppl 1), S80–S89. https://doi.org/10.4103/sja.SJA_203_17

Underwood, L. G., & Teresi, J. A. (2002). The Daily Spiritual Experience Scale: Development, theoretical description, reliability, exploratory factor analysis, and preliminary construct validity using health-related data. *Annals of Behavioural Medicine*, 24 (1), 22–33. https://doi.org/10.1207/S15324796ABM2401_04

Waltz, C. F., Strickland, O. L., & Lenz, E. R. (2010). *Measurement in nursing and health research* (4th ed.). Springer Publishing Company.

Wright, A. G., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment, 31* (12), 1467–1480. <https://doi.org/10.1037/pas0000685>

Yaden, D. B., Kaufman, S. B., Hyde, E., Chirico, A., Gaggioli, A., Zhang, J. W., & Keltner, D. (2022). The development of the Awe Experience Scale (AWE-S): A multifactorial measure for a complex emotion. *The Journal of Positive Psychology, 17*(1), 88–104.

Yoon, S., & Kim, J. H. (2020). Psychometric properties of reverse-coded items in spiritual well-being scales: A cross-cultural validation study. *Mental Health, Religion & Culture, 23*(8), 789–805.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Education in Medicine Journal, 11* (2), 49–54. <https://doi.org/10.21315/eimj2019.11.2.6>

Zamroni, E., Hanurawan, F., Muslihati, Hambali, I., & Hidayah, N. (2022). Trends and research implications of guidance and counselling services in Indonesia from 2010 to 2020: A bibliometric analysis. *SAGE Open, 12* (2), 1–18. <https://doi.org/10.1177/21582440221091261>

Zanon, C., Bastianello, M. R., Pacico, J. C., & Hutz, C. S. (2021). Relationships between positive and negative affect and the five personality factors in a Brazilian sample. *Paidéia (Ribeirão Preto), 31*, e3112.

Zhang, X., & Savalei, V. (2023). Improving the precision of reliability estimates: The role of reverse-coded items in scale construction. *Psychological Methods, 28*(1), 1–18.