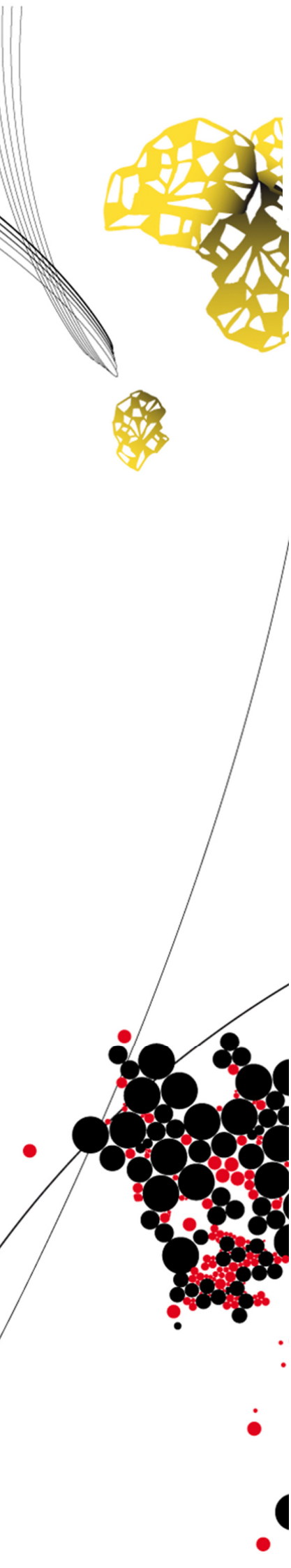

UNIVERSITY OF TWENTE.

Faculty of Behavioural, Management & Social Sciences

Spiritual well-being instruments: A systematic review of the psychometric properties of frequently used scales

Julia M. Pellengahr
s1477536

M.Sc. Thesis
August 2018

Supervisors:
#1 Dr. Vincent van Bruggen
#2 Prof. Dr. Erik Taal
#3 Prof. Dr. Ernst Bohlmeijer

Faculty of Behavioural, Management and Social Sciences
University of Twente
P.O. Box 217

Abstract in English

Introduction: Spiritual well-being (SWB) experienced a constant increase of attention over the past decades, as it correlated beneficially with good mental health and coping strategies especially in difficult life situations. Therefore, SWB might aid the development of clinical treatments to improve people's quality of life. Experts often base the development of interventions and treatments on instruments that measure constructs such as SWB, which requests well-validated and reliable instruments. This literature review, therefore, focuses on the most frequently used SWB instruments of the present age. Accordingly, it investigates the quality of the studies that focused on psychometric properties such as the internal consistency, the test-retest reliability and the content validity. The definition of SWB contains slightly different interpretations among experts. Therefore, one working definition was formulated for this review that was based on two major models that define SWB, namely the 'Four Domain Model' and the 'Two-dimension Model'. **Methods:** A literature review of 459 publications was conducted. This set of references was obtained via two databases: PsycInfo and Web of Science. The most frequently used SWB instruments of the past ten years were extracted. Accordingly, one or more key publications about psychometric properties were determined for each instrument at stake and evaluated by the Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. **Results:** Two most frequently used SWB instruments could be extracted from this review, namely the 'Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale' (FACIT-Sp) and the 'Spiritual Well-Being Scale' (SWBS). According to the COSMIN criteria the quality of those studies ranged from *poor* to *fair*. In conclusion, the statements about the validity and reliability made in those studies need to be treated with caution due to their insufficient quality and more research is needed. **Discussion:** The most important outcome of this review is that the FACIT-SP and the SWBS have only been tested with insufficient quality concerning reliability and content validity yet. The lack of high quality studies concerning psychometric properties might originate from the fact that the SWB definition carries different interpretations and that it is still debated if the religious dimension should be integrated or not within items. Furthermore, many studies administered the instrument only once which was to the detriment of the test-retest reliability. When comparing both instruments, it was concluded that the FACIT-Sp appears more suitable to measure SWB when considering the quality and content of the instrument.

Key words: Spiritual Well-Being, Spiritual Well-Being Scale, Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being Scale, Psychometric Properties

Abstract in Dutch

Introductie: Spiritueel welbevinden (SWB) ondervond de afgelopen decennia een constante toename van aandacht, omdat het positief bleek te correleren met goede geestelijke gezondheids en copingstrategieën, vooral in moeilijke levenssituaties. Daarom zou SWB de ontwikkeling van klinische behandelingen kunnen helpen om de kwaliteit van leven van mensen te verbeteren. Voor de ontwikkeling van interventies en behandelingen is het nodig om constructen, zoals SWB, te meten en dit kan alleen met goed gevalideerde en betrouwbare instrumenten. Dit literatuuronderzoek richt zich daarom op de meest gebruikte SWB-instrumenten van de huidige tijd. Daarom onderzoekt het de kwaliteit van de onderzoeken die zich richten op psychometrische kwaliteiten zoals de interne consistentie, de test-hertest betrouwbaarheid en de inhoudsvaliditeit. Experts verschillen van mening over der correcte definitie van SWB. Daarom werd voor deze beoordeling een werkdefinitie geformuleerd die gebaseerd is op twee belangrijke modellen die SWB definiëren, namelijk het 'Four Domain Model' en het 'Two-dimenison Model'. **Methoden:** Een literatuuronderzoek werd uitgevoerd van 459 referenties die werden verkregen via twee databases: PsycInfo en Web of Science. De meest gebruikte SWB-instrumenten van de afgelopen tien jaar zijn geëxtraheerd. Vervolgens werden een of meer sleutelpublicaties over psychometrische kwaliteiten voor elk instrument in kwestie bepaald en geëvalueerd met behulp van de Consenus gebaseerde Standaarden voor de selectie van medische meetinstrumenten (COSMIN) checklist. **Resultaten:** Uit deze review kunnen twee meest gebruikte SWB-instrumenten worden afgeleid, namelijk de 'Functional Assessment of Chronic Illness Therapy - Spiritual Well-being Scale' (FACIT-Sp) en de 'Spiritual Well-Being Scale' (SWBS). Volgens de COSMIN-criteria varieerde de kwaliteit van die onderzoeken van slecht tot redelijk. Concluderend moeten de uitspraken over de validiteit en betrouwbaarheid, die in die onderzoeken zijn gemaakt, voorzichtig worden behandeld vanwege de onvoldoende kwaliteit en meer onderzoek is nodig. **Discussie:** De belangrijkste uitkomst van deze review is dat de FACIT-Sp en de SWBS met onvoldoende kwaliteit zijn getest op betrouwbaarheid en inhoudsvaliditeit. Het ontbreken van kwalitatief hoogwaardige studies met betrekking tot psychometrische kwaliteit kan voortkomen uit het feit dat de SWB-definitie verschillende interpretaties heeft en dat het nog steeds de vraag is of de religieuze dimensie al dan niet in items moet worden geïntegreerd. Bovendien hebben de meeste studies het instrument slechts eenmaal hebben toegepast, waardoor er geen test-hertest betrouwbaarheid vastgesteld kon worden. Bij het vergelijken van beide instrumenten werd geconcludeerd dat de FACIT-Sp meer geschikt leek om SWB te meten bij het beschouwen van de kwaliteit en inhoud van het instrument.

THE INVESTIGATION OF PSYCHOMETRIC PROPERTIES OF SPIRITUAL WELL-BEING SCALES

Keywords: Spiritueel Welbevinden, Spiritual Well-Being Scale, Functional Assessment of Chronic Illness Therapy-Spiritual Well-Being Scale, Psychometrische Kwaliteiten

Table of Content

Introduction 6
Definition of Spiritual Well-Being (SWB) 6
Beneficial effects of Spiritual Well-Being 8
Operationalization and focus of this review 9
Objective of this review 9
Methods 11
Procedure 11
Inclusion criteria for the dataset 12
Analysis and Evaluation of the Instrument Quality to measure SWB 13
Results 15
Number of Hits and Identified Instruments 15
Introduction to Spiritual Well-Being Instruments 17
Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp) 17
Spiritual Well-Being Scale (SWBS) 19
Discussion 26
Comparing the FACIT-Sp and the SWBS within the scope of this research 29
Strength and limitations of this review 31
Conclusion and directions for further research 33
References 36
Appendix A – The extensive list of the frequencies of all extracted instruments 41
Appendix B – FACIT-Sp Scale (12-item version) 43
Appendix C – SWBS (20 items) 44
Appendix D – Box A, B and D of the COSMIN checklist 45

Introduction

Spirituality positively correlates with a good mental health and indicates less anxiety when dealing with life-threatening diagnoses or traumata; this is what Ellison and Smith (1991), Cotton, Levine, Fitzpatrick, Dold and Targ (1999) and Van Dierendonck (2004) have stated independently from each other since 1991. Furthermore, the chance of suffering depression and experiencing a lower quality of life was suggested to be higher in people with low *spiritual well-being* levels (Chimluang et al., 2017). Over the past years, the attention for spirituality and spiritual well-being has increased steadily as Hawley and Hawley (1993) suggested a trend of more and more people actively searching the purpose of life by trying to find out what it means to be human and living by the inner truth to seek positive attitudes and relationships (Van Dierendonck, 2004). Furthermore, Cobb, Puchalski and Rumbold (2012) also stated that measuring someone's SWB level can be especially helpful within clinical and counselling settings.

Within the research field of health-related issues, constructs are often highly subjective in nature (Mokkink et al., 2010a), thus also spiritual well-being (SWB) can be understood as a subjective experience according to Moberg and Brusek (1978). It is therefore difficult to objectively measure SWB, although the operationalization is highly relevant. Generally, various treatments, interventions, decisions or further research are based on the results of instruments measuring health-related constructs such as SWB. Therefore, well-validated and reliable instruments are a necessity and need to be stressed for future research on topics such as SWB and might open up new possibilities in helping relationships (Mokkink et al., 2010a).

Definition of Spiritual Well-Being (SWB)

It is crucial to mention that literature often uses the constructs of spirituality and spiritual well-being (SWB) interchangeably, as they are closely related to each other (Chowdhury & Fernando, 2013). Simultaneously, it was stated that SWB is known as a consequence and as one operationalization of spirituality and thus spiritual experiences (Monod et al., 2010). In direct comparison, spirituality is a broader theoretical concept that indicates a human condition and covers also aspects such as spiritual beliefs/behaviours/support, whereas SWB is a narrower construct of personal development within the spiritual dimension regarding someone's well-being (Chowdhury & Fernando, 2013). This systematic literature review will mainly focus on SWB, although both terms (spirituality and SWB) will come across as one originates from the other. Generally, all definitions in the field of spirituality are often characterized by diverse interpretations concerning the meaning even among experts (Monod

et al., 2010; Harrington, 2016), therefore a great body of literature was reviewed in order to find the most accepted definition of spirituality and SWB.

SWB interrelates the two constructs *'spirituality'* and *'well-being'* and can be described as an index of one's lived spirituality that indicates one's quality of life in the scope of the spiritual dimension (Chowdhury & Fernando, 2013). It is a subjective experience that comprises the personal search for meaning and purpose in life and the connection with a transcendent dimension of existence (Moberg & Brusek, 1978). Two overlapping models have been presented concerning the elements that form the concept of SWB. On the one hand, the National Interfaith Coalition on Aging (NICA, located in Washington D.C.) concluded already in 1975 that SWB consists of four domains which cohere to determine a person's overall SWB level (Ellison, 1983; Fisher, 2011). A few years later, Fisher agreed with this theoretical model ('four domains model'/'4D model') of SWB (as cited in Gomez & Fisher, 2003). On the other hand, SWB got conceptualized by Ellison with two overall continuous dimensions: the vertical and the horizontal dimension, thus the 'two-dimension model' emerged (Ellison, 1983; Ekşi & Kardaş, 2017). Both models will be described in detail in the following paragraph.

The 'four domains model' summarizes the main themes that were variously mentioned within literature that reviews SWB over the past decades (Fisher, 2011). It is defined as the affirmation of wholeness that is nurtured by the relationship with (1) the self, (2) the community, (3) the environment and (4) God (Fisher, as cited in Gomez & Fisher, 2003; Ekşi & Kardaş, 2017). Firstly, the relationship to oneself belongs to the *personal domain*, which is characterized by one's experience of meaning and purpose, which contributes to a concept of identity and feeling worthy (Gomez & Fisher, 2003). Secondly, the *communal domain* comprises all profound interpersonal relations with significant others that result in love, justice, hope, faith and humanity. Thirdly, the *environmental domain* aims beyond care and nurture for the physical and biological world. A sense of awe, wonder and unity are seen as essential aspects with respect to the environment (Gomez & Fisher, 2003). The fourth relationship forms the *transcendental domain*, which is based on the relationship between the self and some-one/some-thing beyond the human level (such as God, a transcendent reality, a higher existence, a cosmic force etc.). This domain involves faith towards, adoration and worship of the source of mystery of the universe (Fisher, 2011). A balance between all four domains has beneficial effects on one's spiritual health and well-being (Fisher, as cited in Gomez & Fisher, 2003; Fisher, 2011). Besides, the 'two-dimension model' is the second model that has been proposed to define SWB with its components. It incorporates the

religious and existential component: The religious component (vertical dimension) refers to one's relationship with a higher existence or general higher power, thus it is overlapping with the transcendental domain of the 4D model. The existential component (horizontal dimension), in turn, refers to the sense of meaning and purpose in life independent to anything specifically religious, such as the personal domain mentioned above (Moberg & Brusek, 1978). The communal and environmental domains were not explicitly considered by Ellison (Fisher, Francis & Johnson, 2000), however, the existential component is known to incorporate a transcendent link with others and the environment (Riley et al., 1998). Therefore, the existential component could be seen as the integration of the environmental and communal domain of the 4D model as others and environment are addressed as well. For this review a working definition was developed that grasps the construct of SWB and incorporates the overlapping aspects of the above-mentioned models: *SWB is a concept that contributes to someone's holistic health. It can be seen as an acknowledgement of people's spiritual entity by exploring the own purpose and meaning in life, the transcendent interconnection with significant others and the environment as well as finding a relationship with a general higher existence or cosmic force.*

Beneficial effects of Spiritual Well-Being

As already mentioned above, SWB has increased in its importance over the past years and research increasingly focused on this concept (Van Dierendonck, 2004). It received steadily more attention since it was found to positively correlate with one's mental health and general illness prevention, as well as it was found to be beneficial in times of crisis and uncertainties (Peterman, Fitchett, Brady, Hernandez & Cella, 2002; as cited in Van Dierendonck, 2004). Especially when the end of life becomes imminent due to terminal illnesses for example, it is suggested that a high SWB level can function as a buffer against negative feelings (McClain Rosenfeld & Breitbart, 2003) and can facilitate different coping strategies when being in difficult life situations (Ellison & Smith, 1991). It might even result in higher effectiveness of palliative care practices for patients when spiritual needs are addressed as well as existential questions are helped to be answered (McClain et al., 2003). It positively correlates with patients' fighting spirit to persevere (Cotton et al., 1999). Therefore, it can be helpful in understanding the positive effect of certain spiritual beliefs/activities when coping with major stressors (McClain et al., 2003) as well as predicting the quality of life while suffering a serious disease such as cancer or psychopathology (Cotton et al., 1999). Accordingly, experiencing spirituality and a high level of SWB have been found to enhance the feeling of

strength and someone's fighting spirit. Therefore, it would be of high value to detect someone's present level and the effects of SWB and to promote that even more. Instruments, measuring SWB, aid this process as they provide an indication of someone's SWB level. Therefore, further integration needs to be stressed, especially for the operationalization of SWB.

Operationalization and focus of this review

Ellison (1983) described SWB as a continuous variable, which illustrates the assumption that SWB should not be seen as being either present or not, but rather as how much and to what extent it is present within a person. Therefore, the question arises at which SWB level people already are and how to increase this level further. In order to determine someone's level of SWB, a valid and reliable measure of this construct needs to be implemented. According to Peterman et al. (2002) the psychometric properties of widely used instruments to measure spirituality or SWB have not sufficiently been investigated yet.

Therefore, this systematic literature review will focus on the quality of the psychometric properties of the most frequently used instruments that measure SWB in the past decade. As mentioned earlier the interpretations and definitions of spirituality-related constructs highly vary and therefore well-validated measures are needed (Peterman et al., 2002; Harrington, 2016). According to Van Dierendonck (2004) the operationalization of SWB needs to be as direct and plain as possible by investigating people's behaviour, thoughts and feelings, without one's religious orientation as predicting or influencing factor. Thomason and Brody also stressed a high need for exploring the validity of scales that particularly measure spirituality independent from religion to also be suitable for non-religious people who experience spirituality independent form religion (as cited in Peterman et al., 2002).

Objective of this review

The aim of this systematic literature review is therefore to provide an overview of instruments that measure SWB in terms of their psychometric qualities. To limit the scope of this review, it is only focused on the reliability (internal consistency and test-retest reliability) and the content validity of the most frequently used instruments within already existing literature. Therefore, the following research questions will be answered:

(1) "Which are the most frequently used instruments to measure spiritual well-being?"

(2) "What are the psychometric properties of these instruments with regard to reliability?"

(3) “What are the psychometric properties of these instruments with regard to content validity?”

Methods

Procedure

In April 2018, a systematic literature search in two different electronic databases was performed, i.e. PsycINFO and Web of Science (final search date: 20th April of 2018), to identify the most frequently used instruments measuring SWB. Therefore, the following main search terms: "spiritual well-being" OR "spiritual well being" OR 'spiritual wellbeing' were combined with terms concerning psychometric properties, validation and instruments. The final search string was:

Title / Abstract: "spiritual well-being" OR "spiritual well being" OR "spiritual wellbeing"

AND

Title / Abstract: scale OR test OR questionnaire OR assessment OR measure OR inventory OR instrument OR psychometric OR "psychometric properties"

Certain limitations, such as: English Language, Time Frame: 2008-2018 and All Journals/Articles were determined in order to reduce the dataset and make it more specific. The language restriction was based on the fact that the broadest range of references is written in this global language and a first screening revealed that the most of the SWB instruments were originally developed for the English-speaking population. Further, it was chosen to investigate the past ten years of research, thus from 2008 to 2018, as this ensured currently used instruments to be in focus of this review, whereas the body of literature was reduced to a reasonable size. Several combinations were tested to ensure that the dataset includes a spectrum as broad as possible of frequently used instruments that measure SWB. However, a few search terms that seemed logical to be included at first sight, such as religion or spirituality, were not integrated in the final search string due to several reasons that will be explained in detail below.

Historically, the term *religion* was used interchangeably with the term *spirituality* as no actual distinction was made back in time (Morrison-Orton, 2004). However, despite an existing overlap, both terms are nowadays distinguished from each other in the majority of literature (Zinnbauer et al., 1997; Peterman et al., 2002; Harrington, 2016). According to Peterman et al. (2002) both terms became more narrowly defined and therefore even got distinguished recently. For this reason, terms such as "religiosity", "religion" or "religiousness" were avoided. Inclusion of these search terms mainly provided instruments focusing on religion in the first place, without leaving any room for spirituality as it is defined within this review. Furthermore, it became clear that spirituality does not necessarily include

religious beliefs, which means that it can exist on its own (Harrington, 2016). “Spirituality” has also been omitted as an additional search term despite the fact that it is often used interchangeably with SWB, because it is used in various contexts, which would lead to many false positive results utilizing different definitions of spirituality. Thus, the search results got too broad and unspecific by adding spirituality.

A dataset of more than 900 publications was organized by their first author according to the alphabet with the aid of the reference manager “Mendeley” (Version 1.17.13). Due to the scope of the thesis, a systematic random sampling method was applied with an interval of 50 articles. Therefore, the first set of 50 articles was included, while the next set of 50 articles was bypassed, and so on. 500 articles remained and were screened with regard to certain inclusion criteria (see next paragraph). In more detail the titles and abstracts of those articles were screened to discover all used instruments that comprise SWB. Figure 1 (see results section, page 16) illustrates the exact selection procedure of the articles that were used as basis of this systematic literature review. If the quality of the abstract was poor and no information was given about the used measurements, the method sections of those references were taken into account. One half of the dataset was screened by the first author of this review and the other half was done by another researcher writing about a similar topic while eventually using the same dataset. In total, 50 articles were screened in cooperation of both researchers to ensure that the screening process was carried out as similar as possible. The inter-rater reliability coefficient Kappa was .89, which can be interpreted as an almost perfect agreement (Viera & Garrett, 2005). In order to ensure that no fundamental instruments have been left out, a random selection of 50 articles, which did not belong to the dataset of 459 publications, was reviewed.

For every quantitative instrument that included SWB in its name or seemed to operationalize spirituality-related constructs, the frequency of occurrence was noted. Therefore, the most frequently used instruments could be extracted from this list of instruments. Appendix A illustrates the extensive list of the frequencies of all instruments that have been extracted from the dataset.

Inclusion criteria for the dataset

This systematic literature review aimed at identifying the most frequently used instruments to measure the concept of SWB to evaluate the reported content validity and reliability of those instruments. After applying the random sampling method that already reduced the dataset to

500 publications, the titles, abstracts and partly the method sections were screened for the following inclusion criteria:

(1) Integration of the construct of spiritual well-being (with all its different spellings),
(2) At least one quantitative scale was used,
(3) No literature reviews,
(4) SWB-subscales of a broader scale *can* be included,
(5) Articles written in English using SWB-scales in any language,
(6) Scales are explicitly aimed at measuring SWB.

Thus, the 500 publications were screened and the dataset was finally reduced to 459 due to the above-mentioned inclusion criteria. All instruments covering SWB were registered and the frequency of occurrence within this dataset was noted. Next, the five most frequently used scales were examined closely on their items and the overarching construct, to ensure that they indeed did fit within the scope of this literature review.

Analysis and Evaluation of the Instrument Quality to measure SWB

After extracting the most frequently used instruments to measure SWB, additional research was done in order to identify the key publication of each instrument at hand, containing information about the psychometric properties and qualities of the scales. Those publications were accordingly used to evaluate the psychometric properties of the chosen instruments with the aid of the **COSMIN** checklist (**CO**nsensus-based **S**tandards for the selection of health **M**easurement **IN**struments) (Mokkink et al., 2012). For the scope of this review, the evaluation boxes A, B and D (see Appendix D) were used to cover the internal consistency, reliability and content validity, respectively. Each box consists of five to 14 items that need to be answered. This checklist can be used to evaluate the methodological qualities of a study on measurement properties, thus if the investigation of the quality can be rated as reasonable or not (Terwee, Mokkink, Knol, Ostelo, Bouter & de Vet, 2012). However, this checklist does not provide any information about the actual qualities of the instrument itself as it rates the evaluation process that was used to determine the quality of the instrument (Mokkink et al., 2010a).

Table 1 reports the descriptions of the COSMIN checklist and provides the three properties which this review will focus on (Mokkink et al., 2010a). Terwee et al. (2012) developed a scoring system that can be applied to the COSMIN checklist in order to rate each measurement property. Firstly, each item within the evaluation boxes is rated individually as either *poor*, *fair*, *good* or *excellent* (four-point rating scale). Secondly, the overall rating of the

measurement property is determined by extracting the lowest score for any of the items of the box (Terwee et al., 2012). Even though parts of the evaluation might be performed adequately, the overall rating originates from the lowest rating within a box. Appendix D illustrates the list of criteria of the four-point rating scale of each property with detailed information of how it was decided to rate an items as poor, fair, good or excellent.

Table 1

COSMIN Definitions of the Measurement Properties

Domain	***Measurement property***	***Definition***
Reliability	Reliability	The proportion of the total variance in the measurements which is due to 'true' differences between patients
	Internal Consistency	The degree of the interrelatedness among the items
Validity	Content Validity	The degree to which the content of a Health-Related Patient-Reported Outcomes (HR-PROs) instrument is an adequate reflection of the construct to be measured

Results

Number of Hits and Identified Instruments

The above-mentioned search string has yielded a total of 1073 publications, whereas 113 were found to be duplicates. Within the scope of this thesis, the remaining set of 960 publications was too extensive. As described within the method section, this dataset was reduced to a total of 500 articles. Those were then screened for the use of SWB instruments and the inclusion criteria by carefully reading the title and abstract. 41 publications did not fulfil these criteria and the final dataset therefore consisted of 459 publications, the selection procedure is also illustrated in figure 1 in more detail.

The **F**unctional **A**ssessment of **C**hronic **I**llness **T**herapy – **Sp**iritual Well-being Scale (FACIT-Sp) was most frequently used, with a frequency of 153, closely followed by the **S**piritual **W**ell-**B**eing **S**cale (SWBS) which occurred about 129 times. The **S**piritual **W**ell-**B**eing **Q**uestionnaire (SWBQ), the **M**ultidimensional **M**easure of **R**eligiousness and **S**pirituality (MMRS) and the **S**piritual **H**ealth **a**nd **L**ife **O**rientation **M**easure (SHALOM) were used around twelve to 14 times within this dataset. Therefore, it was chosen to focus this review on the FACIT-Sp and the SWBS according to their exceptionally high frequencies in comparison to the other scales.

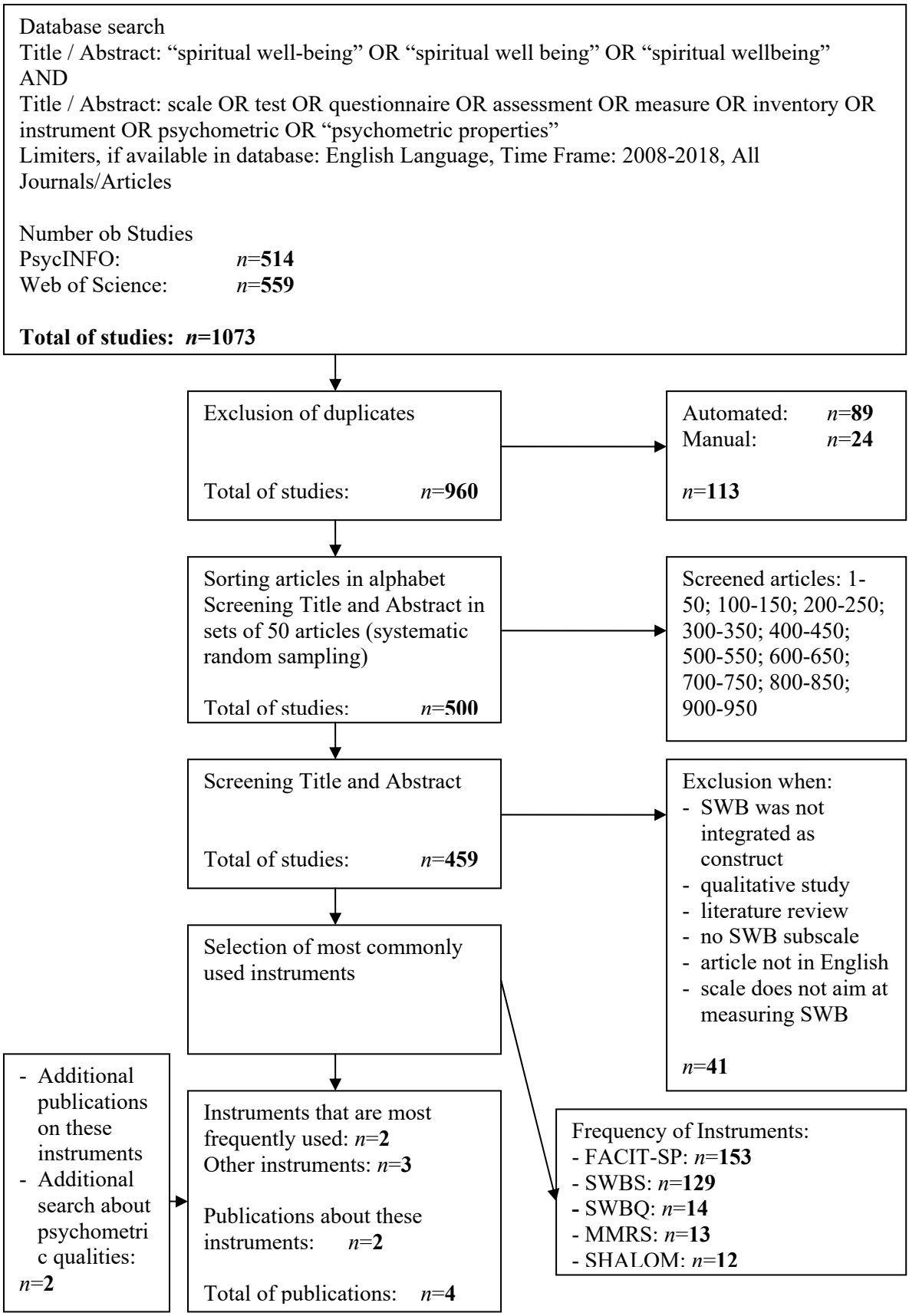

Figure 1. Summary of the exact selection procedure of the data.

Introduction to Spiritual Well-Being Instruments

In the following paragraphs the two most frequently used SWB instruments will be described in more detail, thus the Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp) and the Spiritual Well-Being Scale (SWBS). The key publications of each instrument will be evaluated in terms of the quality of how they investigate the psychometric properties of the scales. The focus lies on the reliability (internal consistency and test-retest reliability) and the content validity. Table 2 (see page 22 onwards) provides a summary of the general psychometric characteristics and the quality and ratings of the key publications. In case that the key publication did not cover all of the psychometric properties at stake, it was systematically researched for other articles to not overlook anything.

Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp)

The FACIT-Sp is a measure originating from the larger FACIT measurement system, which consists of one core instrument (FACIT-General) and over 50 different FACIT scales and symptom indexes (Peterman et al., 2002). The website of the FACIT-Systems (http://www.facit.org/) provides information about the FACIT-Sp, which belongs to the category of the non-cancer specific measures. The FACIT-G investigates someone's general quality of life and was developed by analysing and coding interviews of 135 cancer patients and 15 oncology specialists. Consequently, the inclusion of SWB for the general quality of life got stressed by those interviews as spiritual concerns turned out to be comforting. Due to low factor loadings of the SWB items within the FACIT-G scale, it was decided to develop a separate instrument concentrating on the spiritual dimension (Peterman et al., 2002). Therefore, 200 interviews were conducted with psychotherapists, religious/spiritual experts as well as cancer patients during this developmental process and additionally some items were adopted from the FACIT-G (Peterman et al., 2002).

As a consequence, the FACIT-Sp was developed with its five subscales: physical well-being (seven items), social/family well-being (seven items), emotional well-being (six items), functional well-being (seven items) and spiritual well-being (twelve items), as those constructs all contribute to someone's quality of lfie. Researchers that solely focus on SWB within their studies, commonly used the FACIT-Sp-12 with its twelve items comprising only the spiritual dimension of well-being (see Appendix B). Those twelve items can be answered by a five-point Likert Scale (0 = "Not at all" to 4 = "Very much") and those statements can be subdivided into two subscales. The first subscale is Meaning/Peace (eight items) with items

such as “I have a reason for living”. Secondly, there is the Faith subscale (four items) consisting of items such as “I find strength in my faith or spiritual beliefs”. Peterman et al. published the first official investigation and demonstration of the psychometric properties of the FACIT-Sp in the year 2002, therefore this reference has been selected as key publication. Within this publication, two studies were conducted to demonstrate the reliability and to examine the validity of the FACIT-Sp in comparison to five other measures. For the first study, 1617 cancer patients from the United States of America and Puerto Rico participated, whereas the second study included 131 participants of whom 65% reported an affiliation with a religious tradition (Peterman et al., 2002). In this article information is given about the internal consistency by calculating Cronbach’s alpha, nevertheless it is unclear how missing items were handled and if the sample size is adequate (Cronbach’s alpha: Total scale: 0.87, Meaning/Peace: 0.81, Faith: 0.88). The FACIT-Sp was further found to be positively correlated to the FACIT-G, the **P**rofile **o**f **M**ood **S**tates (POMS) and the **M**arlowe-**C**rowne **S**ocial **D**esirability **S**cale (MCSDS) which might support the instruments validity (Peterman et al., 2002). The subscale Meaning/Peace was notably stronger correlated with its eight items than the Faith subscale, however, in conclusion Peterman et al. found the FACIT-Sp to be valid and reliable to measure SWB in a sample of chronically ill people (Peterman et al., 2002). However, all those findings need to be treated carefully as long as the quality of the study is unclear.

Evaluation: Firstly, the evaluation of the reliability (internal consistency and test-retest reliability) of the FACIT-Sp was assessed with the aid of the COSMIN checklist, see table 2 for detailed results. The above-mentioned key publication from Peterman et al. (2002) was the first publication that focused on the psychometric properties of the FACIT-Sp, however, the test-retest reliability was not addressed at all as only one administrations took place. Additional literature apart from the data set was systematically screened for test-retest reliability; however, no publications were found. Therefore, the evaluation of the test-retest reliability had to be rated as *poor*. Secondly, the evaluation of the internal consistency could be rated as *fair*, since no statement was made about how missing items were handled. Furthermore, Peterman et al. (2002) reported in this key publication a good internal consistency (Cronbach’s α = .87) regarding the entire scale and calculated separate alphas for the two subscales (Cronbach’s α = .81 and .88). Calculating alpha is only suitable for unidimensional (sub-)scales and therefore these findings need to be treated with caution due to the questionable quality and adequacy of the investigation of the internal consistency.

Thirdly, the evaluation of the content validity was put into focus and will be stated in the following. The twelve items of the FACIT-Sp were constructed in cooperation with cancer patients, psychotherapists and experts within the field of spirituality and religiousness. Thus, the original target population of the FACIT-Sp (cancer patients) was deeply involved in the development process of the items by conducting and analysing interviews, whereas several experts in the field of spirituality ensured that all important factors and aspects were pointed out. A clear description of the measurements aim is provided, however, the process of deciding which items will be included was not clearly stated within this study. Furthermore, it was not investigated if the items comprehensively reflect the construct to be measured. Therefore, the evaluation of the content validity had to be rated as *poor*.

In conclusion, the COSMIN criteria revealed the evaluation of those psychometric properties to be rather *poor* concerning the test-retest reliability and the content validity or *fair* concerning the internal consistency. Therefore, the statement of Peterman et al. (2002) about the FACIT-Sp being a reliable and valid instrument needs to be treated with caution. The overall quality of the study of Peterman et al. (2002) is not adequate to make such strong statements about the content validity and the reliability.

Spiritual Well-Being Scale (SWBS)

Paloutzian and Ellison (1982) developed the SWBS, which is known as one of the most widely used instruments to measure SWB today (see Appendix C). It consists of 20 items and the respondents are asked to indicate on a six-point Likert scale to which degree they agree/disagree with religious and existential statements (SA = “Strongly Agree” to SD = “Strongly Disagree”). The SWBS can be split into two subscales: (1) The Religious Well-Being (RWB) subscale (ten items, α = .87), which aims to investigate the degree to which a relationship to a higher existence (God) is experienced as pleasant and satisfying (Paloutzian & Ellison, 1982). An example of one item is: “I believe that God loves me and cares about me.” (2) The Existential Well-Being (EWB) subscale (ten items, α = .78), which refers to the sense of satisfaction and purpose in life. An example of one item is: “I feel very fulfilled and satisfied with life.” According to Paloutzian and Ellison (1982) half of the items need to be scored reverse, as ten items are negatively and ten items are positively formulated to prevent response set problems. This scale is known as the first instrument assessing SWB in two dimensions: existential and religious dimension (Paloutzian & Ellison, 1982).

Genia (2001) published a study which focuses on the psychometric properties of the SWBS. Several authors concentrated on psychometric properties even before Genia did in

2001, however, those articles were either not accessible within the scope of this thesis or were aimed at an extremely specific group of respondents. Therefore, the publication of Genia (2001) was chosen as a key publication of the SWBS. This study was aimed at pointing out the significant psychometric problems of the scale and to investigate those in more detail. 211 college students, who identify with a broad spectrum of different religious traditions or are "non-traditionally religious", participated in this study. Hence, non-religious students have not participated at all. Genia (2001) reported a high internal consistency of the SWBS (Cronbach's Alpha = .76 - .94, see table 2) without stating how missing items were handled. Furthermore, a ceiling effect due to certain religious affiliations and a high skewness appeared to be problematic in this article. The SWBS also seems to include more than the two existing factors. As Genia (2001) generally does not provide any information about the test-retest reliability and content validity, a second reference was consulted to cover those aspects: the reference of Soleimani et al. (2017). This article, however, focused on psychometric properties of the *Persian Version* of the SWBS, which was developed by forward-backward translation of experts. According to Soleimani et al. (2017) the Persian Version of the SWBS was found to be a reliable and valid instrument, while the two underlying factors of the scale could be confirmed: the relation with God and the relation with life. However, further factor analysis revealed the presence of a third factor construct within the SWBS, which might be attributable to the Iranian culture that dominated the sample (Soleimani et al., 2017). The significance of the above-mentioned coefficients and results need to be treated with caution until the quality of the study is investigated.

Evaluation: The SWBS is based on the two-dimensional model mentioned above and therefore the RWB and EWB subscales cover the transcendental and existential domain of the 4D model, respectively. The authors of the scale, Paloutzian and Ellison (1982), claimed that the total scale has a reasonable test-retest reliability (r = .89), however, this article was not chosen to be the key publication as it missed detailed information about reliability and content validity. Genia (2001) mainly concentrated on psychometric problems of the SWBS and therefore did not investigate all the psychometric properties that are in focus of this review. The internal consistency was calculated with fair quality due to missing information on how missing items were handled. The test-retest reliability and the content validity were not addressed at all. Generally, there was a lack of literature regarding the exploration of the psychometric qualities of the English version of the SWBS. However, Soleimani et al. (2017) thoroughly investigated the reliability and validity of the Persian Version of the SWBS with the aid of qualitative and quantitative methods. The test-retest reliability was accurately

assessed by the two-way mixed intraclass correlation coefficient (ICC, “almost perfect” when > .81) with a two-week interval. The ICC of the whole scale resulted to be .825 and could be rated as almost perfect for the SWBS within a 95% confidence interval ranging from .795 to .853 (Soleimani et al., 2017). However, it was not stated how missing items were handled, which only leads to a *fair* quality for the evaluation of the test-retest reliability. Soleimani et al. (2017) calculated the coefficient alpha for the internal consistency of the two subscales (RWB: α = .92; EWB: α = .84), which was interpreted as good. However, the quality of this calculation was again only *fair* as missing items were still not mentioned at all.

In order to evaluate the content validity of the SWBS, qualitative and quantitative methods were conducted. A team of 15 experts was involved to assess the wordings, scaling and item allocations qualitatively. Furthermore, those experts calculated the Content Validity Ratio (CVR, “appropriate” when > .49) and the Content Validity Index (CVI, “appropriate” when > .79) with a three- and four-point scale, respectively (not essential, useful but not essential, essential) (Soleimani et al., 2017). No item had to be excluded with regard to the content validity. However, it was not investigated to what extent the items suit the target- or sample population. Therefore, the process of investigating the content validity needs to be rated as *poor*, even though the rest of investigation was performed adequately.

In conclusion, the Persian version of the SWBS seemed to be reliable and to show a good content validity by Soleimani et al. (2017). However, those findings need to be treated carefully due to two reasons. First of all, the COSMIN evaluation revealed a *poor* quality concerning the content validity, whereas the internal consistency and the test-retest reliability were investigated with a *fair* quality. Therefore, it appears again to be questionable if this instrument is reliable or valid such as Soleimani et al. (2017) and Genia (2001) have suggested due to the lacking quality of their studies. Second of all, the evaluation of the English version might bring different results due to cultural differences, translation issues or differences within religious traditions among other cultures.

Table 2

Psychometric Characteristics and Quality of the Instruments

	Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp)	***Spiritual Well-Being Scale (SWBS)***	
Key Publication	Peterman, Fitchett, Brady, Pharm and Cella (2002)	Genia (2001)	Soleimani, Sharif, Allen, Yaghoobzadeh, Nia and Gorgulu (2017) – Persian Version of the SWBS
Total number of Items	12	20	20
Scaling	Five-point Likert Scale ("not at all" – "very much")	Six-point Likert Scale ("Strongly Agree" – "Strongly Disagree")	Six-point Likert Scale ("Strongly Agree" – "Strongly Disagree")
Subscales of the construct	Spiritual well-being: - peace/meaning (8 items) - faith (4 items)	Spiritual well-being: - religious well-being (10 items) - existential well-being (10 items)	Spiritual well-being: - religious well-being (10 items) existential well-being (10 items)
Study Population	1617 predominantly cancer patients from the USA and Puerto Rico	211 college students, 34% were Catholic, 29% were Protestant, 13% were Jewish and 24% were unaffiliated to faiths	300 adult outpatients of primary care clinic sites from Iran
Length of follow-up	N/A	N/A	2 weeks
Dropouts + missing items and the way they were handled	-	-	-

	Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp)	***Spiritual Well-Being Scale (SWBS)***	
Reliability			
Test-Retest Reliability	-	-	Test-retest reliability: ICC = .825
Study quality (conform COSMIN)	Poor (only one measurement)	Poor (only one measurement)	Fair (not clear how missing items were handled)
Internal Consistency	Coefficient Alpha FACIT-Sp: α = .87; Meaning/Peace subscale: α = .81; Faith subscale: α = .88	Coefficient Alpha of the total SWBS: - Catholic: α = .93 - Protestant: α = .91 - Jewish: α = .76 - non-traditionally religious people: α = .88 RWB: - Catholic: α = .94 - Protestant: α = .93 - Jewish: α = .91 - non-traditionally religious people: α = .93 EWB: - Catholic: α = .91 - Protestant: α = .78 - Jewish: α = .84 - non-traditionally religious people: α = .87	Coefficient Alpha of the RWB: α = .92 EWB: α = .84

	Functional Assessment of Chronic Illness Therapy – Spiritual Well-being Scale (FACIT-Sp)	***Spiritual Well-Being Scale (SWBS)***	
Study quality (conform COSMIN)	Fair (not clear how missing items were handled)	Fair (not clear how missing items were handled)	Fair (not clear how missing items were handled)
Validity			
Content validity	The items of the FACIT-Sp were constructed from analysing interviews with psychotherapists, cancer patients and religious/spiritual experts.	-	15 experts were consulted to assess and comment the items itself, the wording, the allocation and the scaling.
Study quality (conform COSMIN)	Poor (not assessed if all items together comprehen-sively reflect the construct to be measured)	Poor (not assessed if all items refer to relevant aspects of the construct to be measured, not assessed if all items are relevant for the study population or target population involved, not assessed if all items together comprehensively reflect the construct to be measured)	Poor (not assessed if all items are relevant for the study population or target population involved)
Overall evaluation of	Poor: 2 Fair: 1	Poor: 2 Fair: 1	Poor: 1 Fair: 2

the quality of the primary publication. Sum of given qualifications.	Good: N/A Excellent: N/A	Good: N/A Excellent: N/A	Good: N/A Excellent: N/A

Note. N/A = Not Applicable and - = Not Available

Discussion

This review was aimed at answering three research questions addressing the most frequently used SWB instruments in terms of their reliability and content validity. Accordingly, two instruments that measure SWB were extracted from the used dataset and were thoroughly evaluated with the aid of the COSMIN Checklist. Thus, the first research question: (1) "Which are the most frequently used instruments to measure spiritual well-being?" was answered by extracting the FACIT-Sp and the SWBS. As this review focused on the past ten years of research, it can be concluded that the FACIT-Sp and the SWBS are current key instruments to measure SWB. As indicated in Appendix A, a long list of diverse instruments that measure SWB or spirituality-related constructs emerged from the dataset of this review. The majority of scales existed only once or twice within the dataset, however, the variety of different scales might be attributable to the various existing interpretations of SWB definitions (Harrington, 2016). The FACIT-Sp and SWBS, however, occurred exceptionally often and became therefore the focus of this review.

The second and third research questions were answered by evaluating the quality with which the psychometric properties were assessed: (2) "What are the psychometric properties of these instruments with regard to reliability?" and (3) "What are the psychometric properties of these instruments with regard to content validity?" The quality of the study that investigated the test-retest reliability was evaluated as *poor* for the FACIT-Sp and as *fair* concerning the SWBS, meaning that no actual statement can be made about the consistency of these scales over time. The internal consistency of a (sub-) scale states to what extent the items are interrelated and produce similar scores; however, also concerning this aspect the quality of the studies was only *fair*. Hence, it is not possible to affirm the internal consistency that was suggested by the key publications. The quality of studies in which the content validity was evaluated was *poor*, therefore not much can be said about the fit of the items and the extent to which the two scales cover all facets and aspects of SWB. The SWBS and the FACIT-Sp, however, are both famous instruments that are frequently used recently, which brings up the question; why till now the psychometric properties of those instruments have not been evaluated in a more thorough way. In the following paragraph, several possible reasons will be outlined.

Most notable is that the content validity of both instruments has been rated as poor since at least one item within the COSMIN checklist had to be rated as poor. The development of the items of the FACIT-Sp took place in cooperation with experts and actual end users of the scale, whereas for the SWBS the developmental process of the items is

unclear. However, for both instruments it was not thoroughly stated how it was decided which items are relevant for the measurement instrument, which eventually lead to a lower rating of the overall quality of content validity. The content validity is defined as a matter of judgement that consists of two phases. Firstly, a careful conceptualization of the construct is needed before items are developed and secondly, assessments by experts are necessary to ensure good quality (Polit & Beck, 2006). Polit and Beck (2006) suggested a higher need for clarity and effort in investigating content validity in general. Even among experts, SWB has shown to be a construct that has different interpretations of definitions (Harrington, 2016), which might have caused difficulties in investigating the content validity of SWB instruments. Therefore, the content validity of instruments might be problematic as several empirical definitions of SWB have been proposed and used. Especially, since the debate continues if spirituality and religion should be used interchangeably (Chowdhury & Fernando, 2013), whereas other researchers state a high necessity of distinguishing both constructs and to develop SWB instruments that are not influenced by certain religious traditions (Van Dierendonck, 2004; Chirico, 2016). This might explain the poor rating of content validity established by the review.

Additionally, the review was based on the COSMIN checklist, a well-validated instrument, to judge the quality of studies concerning measurement properties within the health-related spectrum (Mokkink et al., 2010b). According to Mokkink et al. (2010b) adequate training and some experience is needed in order to adequately evaluate instruments with the aid of the COSMIN criteria. However, due to the fact that this review is written within the scope of a Masterthesis, the missing trainings needed to be compensated by thoroughly studying and following the COSMIN manuals. A detailed description (see Appendix D) of how to rate the items was provided and utilized thoroughly, which decreased the chance of false ratings.

Another central aspect which might have affected the quality ratings of the COSMIN checklist, is that within the key publications it was not always apparent if certain methodological aspects were solely not appropriately reported or not even performed at all. Thus, clear evidence was missing if the reporting or the methodological quality was actually poor and it still can be assumed that they key publications were properly executed whereas the description of all actions was insufficient.

Besides this, the COSMIN ratings are rather strict as the worst rating within a box determines the overall rating for the entire box. Thus, for example, the internal consistency needs to be rated as fair as soon as it is not clearly stated how missing items were handled,

even though the rest of the criteria might be performed with excellent quality. It might be assumed that that this item might only be of fair quality due to the above-mentioned insufficient reporting, but still it would be rated with fair quality.

When reviewing the literature, it was noticeable that most articles include a proper investigation of the internal consistency with the aid of Cronbach's alpha as it is the most widely used reliability coefficient (Streiner, 2003). Further, most studies only administered the instrument once, which in turn causes a lack of literature considering the test-retest reliability, which had already been stated by Monod et al. (2010) as well.

Besides this literature review on SWB instruments, another review by Monod et al. (2010) exists which mainly concentrates on instruments of the broad construct of spirituality and more narrow constructs such as SWB, spiritual needs or spiritual coping. Due to the above-mentioned exclusion criteria, no literature reviews were included in the dataset. Apart form the dataset, that reference was therefore extracted to discuss and compare the quality and results of both reviews. Monod et al. (2010) mainly support the above-mentioned findings, such as the FACIT-Sp and the SWBS being the most frequently used instruments to measure SWB, as well as a lack of literature concerning some of the psychometric properties. The scoring system that was used to rate the quality of the scales by Monod et al. (2010) was specifically developed for that study and it was rather simple in comparison to the COSMIN checklist. For each property one or two items were formulated that could be either rated with one or zero points, and the maximum score is a six. The SWBS got rated with a six, whereas the FACIT-Sp got rated with a five by Monod et al. (2010). This differs strongly from the COSMIN ratings that both scales achieved (poor to fair). The evaluation criteria were from different nature, as the COSMIN criteria consist of more items per property, consider more detailed aspects of each property and are created by experts through conducting an international Delphi study. Whereas the scoring systems of Monod et al. (2010) was developed for that review based on recognized standards in general instrument development and was kept simplistic with one or two general items for each property. One example to assess the reliability was formulated as "Internal consistency (Cronbach Alpha)", which could then be answered with "1" or "0", depending on if it was calculated or not. This might therefore explain the gap between the ratings among both reviews and strengthen the quality of this review as the well-validated COSMIN criteria were used to rate the quality of evaluation.

Comparing the FACIT-Sp and the SWBS within the scope of this research

After closely reviewing the FACIT-Sp and the SWBS major differences became obvious and were identified within this review. Both scales are aimed at measuring the same construct of SWB, however, possible differences might lead to a conclusion that one scale might be more suitable in measuring SWB, also within the working definition of this review, as the other.

The SWBS (Paloutzian & Ellison, 1982) consists of 20 items in total, whereas ten items include the word "God" (RWB-subscale). This emphasizes strongly the affiliation with theological traditions that require the existence of a God (Fisher et al., 2000). When being affiliated with a religious tradition that focuses on a general higher existence and not particularly on a God, respondents might be irritated. Peterman et al. (2002) stated additionally that the focus on a close relationship to God characterizes the Evangelical Protestantism, whereas this relationship is less significant and relevant within other faith groups. Research has already illustrated problematic ceiling effects within evangelical Protestantism samples, which supports the above-mentioned critic even more (Bufford et al., 1991; Genia, 2001). Therefore, the development of SWB instruments, that are independent from specific religious traditions, is highly needed (Thomason & Brody, as cited in Peterman et al., 2002; Dierendonck, 2004), especially since spirituality can exist without being particularly religious (Chirico, 2016; Harrington, 2016). A possible solution to decouple religious affiliations from the SWBS within non-religious samples might be, to only administer the EWB subscale which is free from any religious themes.

In contrast, it can be mentioned that the two models thoroughly described in the introduction section of this review explicitly state that the existence of God or a general higher existence forms a crucial implication for SWB. It is furthermore important to state that a study by Genia (2001) reported Cronbach's Alpha values ranging from good to even excellent among different religious affiliations (Catholic, Protestant, Jewish and even non-traditionally religious people). Therefore, according to Genia (2001) the SWBS is applicable for universal administration among different religious traditions, except from non-religious people who have not been part of this study. Bufford et al. already expected in 1991 that the EWB items would show similar results in religious as well as non-religious samples, which would suggest that this subscale (EWB) is not influenced by religion as it does not include "God" particularly in contrast to the RWB subscale. However, those findings and suggestions need to be treated with caution since properly administered research to validate the EWB for non-religious people is still missing to this point in time. It would therefore be valuable to further investigate on this topic, namely which role religious affiliations play in the definition

of SWB and therefore the development in SWB instruments in the future. It might even be an improvement of the SWBS to substitute the word "God" with a higher existence, in order to address all different kinds of affiliations. The EWB subscale of the SWBS in comparison is clearly independent from any religious beliefs and traditions as it focuses on someone's general life satisfaction and purpose in life as already mentioned above.

The FACIT-Sp (Peterman et al., 2002) remains free from any specific identification with religious affiliations as the instrument focuses generally on faith, spiritual beliefs and life satisfaction (Faith-subscale) and on the purpose in life, feeling peaceful and comforted (Meaning/Peace-subscale). This instrument crosses therefore a broad range of religious and non-religious traditions, so that even non-religious respondents can identify themselves as being spiritual without being religious (Peterman et al., 2002). In comparison, the SWBS might be less generalizable and more influenced by religious affiliations of respondents, whereas the FACIT-Sp might be less generalizable due to its focus on chronically ill people. This aspect will be further illustrated below.

The generalizability issue of the FACIT-Sp leads to the important aspect that both instruments were originally aimed at different populations. The FACIT-Sp is originally developed for people suffering from a chronic illness and was therefore created in cooperation with cancer patients. The generalizability might be not applicable for patients without any chronic illness. However, meanwhile some scales from the FACIT-Systems have also been validated within populations that did not suffer any chronic illness. This might also be applicable to the FACIT-Sp after thorough investigation with different population groups (Cella et al., 2018). In contrast, the SWBS was not specifically developed for a certain population, although the authors recommended using this instrument for clinical and counselling purposes (Ellison, 1983). Therefore, the SWBS was originally created for a broader population than the FACIT-Sp.

Next, the key publication which investigated the content validity of the SWBS mainly focused on the Persian version of the instrument instead of the original English version. Therefore, this publication based its results therefore on a non-English speaking population. According to Ekşi and Kardaş (2017) spirituality-related constructs can hardly be disengaged from the cultural context in which they were developed, especially when certain religious beliefs are linked to the cultural context. Therefore, it is questionable if this instrument is applicable within other cultures than the English culture with different religious beliefs (apart from Christianity). Generally, the translated version is not guaranteed to have the same content validity as the original version (Brislin, 1970). This is an issue that might be

decreased by a proper forward-backward translation (McDermott & Palchanes, 1992), which Peterman et al. (2002) performed to increase cross-cultural validity of the instrument, even though content validity stays always sensitive to cultural aspects. According to this argumentation the content validity of the FACIT-Sp is expected to not have been influenced by cultural factors. The FACIT-Sp, however, was translated and validated in over 50 different languages by experienced experts (Cella et al., 2018). Due to its focus on general life satisfaction and purpose in life, the FACIT-Sp is expected to be less sensitive for biases resulting from religious traditions and beliefs.

Considering all these above-mentioned findings and according to the scope and aspects of this review, the FACIT-Sp seems to be more suitable to measure SWB in comparison to the SWBS. Apart from not knowing if religion should be taken into consideration when measuring SWB or not, the FACIT-Sp measures SWB anyway without the influence of religious beliefs and affiliations and can therefore be used more universal among different religious traditions. Generally, both instruments are obviously based on slightly different definitions of SWB as one integrates religious beliefs whereas the other is free from this dimension. Therefore, the SWBS might be restricted to the religious population that is affiliated to "God". The main target population of the FACIT-Sp are chronically ill people. Thus, the FACIT-Sp needs more investigation in terms of the generalizability among a general population. The FACIT-Sp is further a scale that has already been translated and validated in over 50 different languages and is due to its non-religious focus expected to show cross-cultural validity. Overall, the FACIT-Sp seems to be the more suitable instrument to measure SWB due to its generalizability, the well-validated translations, the non-religious orientation and the consultation of experts during the developmental process.

Strength and limitations of this review

The first and most noticeable strength of this review is that, as far as known, no systematic evaluation of the most commonly used SWB instruments with the aid of the COSMIN checklist had been performed yet. The COSMIN criteria are detailed and are rather strict as the worst item-rating decreases the quality of the whole psychometric property, however it is well-validated and was thoroughly developed by experts administering a international Delphi study and it focuses especially on health-related patient reported outcomes (HR-PROs). Therefore, the COSMIN checklist forms a strong basis in order to appropriately rate the quality of the key publications. Furthermore, this review solely focuses on SWB instruments rather than the multidimensional construct of spirituality that also includes other aspects such

as SWB, spiritual coping or spiritual needs (Monod et al., 2010). Therefore, another strength point of this review is the narrow focus on SWB, as Monod et al. (2010) criticized its broad focus on spirituality that also includes other dimensions apart from SWB. However, this review also has limitations regarding the search string, the instruments and the dataset, which will be described in more detail below.

First of all, it has to be mentioned that the search string was limited and that the actual dataset of 960 articles had to be reduced to its half with the aid of a random sampling method, due to the scope of this review. Those aspects might have led to missing relevant instruments. However, the chance of missing relevant instruments has been reduced as much as possible by checking the measurement sections of the databases, where all instruments regarding the overarching search terms are presented. Furthermore, the combination of applying the random sampling method on an alphabetically ordered dataset (by the first author) might have caused the exclusion of certain authors. The search string was held simple with "spiritual well-being" (in all three different spellings) as one of the main search terms. While determining the final search string, it became obvious that there is no uniform definition for spirituality-related constructs (Monod et al., 2010; Harrington, 2016). Due to that, spirituality or SWB are not clearly differentiated from religion and adding constructs such as "religion" broadened the dataset enormously while including many irrelevant instruments. Thus, due to the indistinct definition and the scope of this thesis, it was decided to reduce the search string to "spiritual well-being" as main search term in combination with methodological search terms. Additionally, the two most frequently used instruments occurred over 120 times, whereas others occurred only between one or 14 times. Thus, it can be stated with confidence that the relevant instruments have been included due to the massive difference between the frequencies.

Second of all, it is important to mention that the FACIT-Sp originally emerged from the FACIT-G and was especially developed to include the spiritual dimension when measuring someone's quality of life, as already described above. However, next to the FACIT-Sp, the FACIT-Sp-12 is often used depending on the focus of the study as already described above (Cella et al., 2018). Thus, either the FACIT-Sp or the FACIT-Sp-12 Version was used within the articles of the dataset. As the abstracts and titles did not always indicate whether the twelve-item version or the full version of the FACIT-Sp was used, no difference was made during the process of extracting the frequency of how often each scale was used. Hence, the frequencies of the FACIT-Sp might be over estimated as both versions were totalled up. However, it was ensured that the key publication focused on the twelve-item

version as those items indeed address SWB (Peterman et al., 2002). The reason why this is mentioned within the limitations section is that the exact frequencies of each FACIT-Sp(-12) version can not be identified.

In order to apply the systematic random sampling method as the dataset had to be reduced to its half, all references were organized by their first authors according to the alphabet. This was done in order to be able to manually detect the remaining duplicates that had not been identified by the used reference management software package, called Mendeley. Furthermore, an alphabetic order seemed to be most systematic to get a fast overview of key authors for further research. However, the combination of this sorting strategy and applying the above described random sampling method (interval of 50), might have caused that the extracted frequencies of instruments are biased as some authors published several articles using the same instruments. Therefore, the frequencies of the instruments must be treated with caution. However, since the two most commonly used instruments (FACIT-Sp and SWBS) occurred excessively often in comparison to less frequently used scales, this aspect has probably not made such a difference. Additional 50 articles apart from the final dataset were screened for further instruments as well as the database features that give an overview of the frequencies of all instruments that occur with this particular search string. Those actions were carried out to prevent a biased dataset as much as possible; however, no further instruments had to be extracted either from the additional articles or from the screening of the database features.

Conclusion and directions for further research

To date, the most frequently used instruments have not adequately been tested on their psychometric qualities with regard to the internal consistency, the test-retest reliability and the content validity as far as this review suggests. Apart from the reviewed articles it would be of high value to focus thoroughly on the investigation of the psychometric properties and the generalizability of the FACIT-Sp and the SWBS. Within the scope of the reviewed articles, the procedure of evaluating these properties was rated ranging from *poor* to *fair* according to the COSMIN checklist. This finding emphasizes that the statements of the key publications, that the instruments were actually found to be reliable and valid, should be handled with caution. It is therefore to be concluded that the actual reliability and content validity of the scales could not be extracted due to the fair/poor quality in the evaluation process of those properties. Although there is a lack of evidence for good quality, these instruments are frequently used, as was also illustrated by this review. This review led further to the

conclusion that the FACIT-Sp seems to be more suitable when measuring SWB than the SWBS. This decision was based on the findings that the FACIT-Sp seems more generalizable due to the independency to religion, the well-validated translations and the consultation of experts during development. The working definition of this review includes the existence of a higher existence or higher powers as this got stressed by the two main models mentioned in the introduction. However, the SWBS specifies the higher existence to be a "God" that does not suit the diversity of religious affiliations and excludes non-religious people right away.

Therefore, an important aspect in the further development of SWB instruments will be the formulation of a more universal and clear definition of SWB that can be used among various cultures and religious/non-religious traditions. Accordingly, the debate whether religion should be integrated in the definition of SWB or not could therefore be resolved. This would help the investigation of the content validity and provide an overview of the instruments that appropriately measure SWB. It might already be enough to substitute the word "God" of the RWB with more general terms such as a higher existence or a cosmic force as it was done within the working definition of this review. However, more research is needed on that topic.

As the religious dimension seems to restrict the generalizability of the SWBS, a recommendation for further research might be to focus on the non-religious EWB subscale of the SWBS and to reformulate the RWB subscale with general terms that suit religious as well as non-religious people. Especially, within populations that follow various religious beliefs those adjustments might prevent the above-mentioned ceiling effect or other biases that might be caused by religious affiliations and the utilization of the term "God". Accordingly, it might be of high value to investigate the generalizability of the EWB subscale among different religious and non-religious groups. As stated above someone can be spiritual without being affiliated to a particular religious tradition; therefore, one SWB instruments is needed that is not influenced by religious aspects. That fundamental change of focus to decouple SWB scales from religious traditions might be an aim for further research as well, which was already suggested by Thomason and Brody in 1999 (as cited in Peterman et al., 2002). Another recommendation for further research is to aim at more diversity when using the original total SWBS. The religious traditions should be controlled, so it only aims t homogenous samples with a similar religious orientation. This would not eliminate the influence of religious affiliations, however, as long as all respondents follow similar beliefs there is a high chance that it is balanced out. Even though it cannot be compared to other studies then, it would at least provide an overview of the used sample.

Furthermore, the literature that properly evaluates the content validity and reliability of instruments was lacking within the scope of this review. It would therefore be of high value to conduct additional literature reviews specifically aiming at that issue and to administer more research on the quality of both instruments that fulfil the COSMIN checklist in terms of the reliability and the content validity. Accessible literature regarding the psychometric properties of the SWBS and the FACIT-Sp was generally hard to find.

For further research it can therefore be suggested to use the COSMIN checklist as guideline when planning, administering and reporting the study. The quality of the research will certainly increase through the detailed description within the COSMIN boxes what steps need to be taken when appropriately investigating the content validity for example. Besides, the taken steps, argumentations and findings need to be documented fully transparent, so that the study is replicable and that other researchers can comprehend each single step. Furthermore, it is highly important to use a universal system to evaluate psychometric properties to avoid different ratings of the same scale among researchers. Monod et al. (2010) for example rated the SWBS and the FACIT-Sp differently in comparison to the results of this review, which might be caused by the different scoring systems. Additionally, researchers using the COSMIN checklist to rate scales, should be trained sufficiently to ensure adequate and objective administration. To increase the objectivity even more, several raters should be responsible to judge the quality of the instrument at hand, so that an inter-rater reliability can be calculated.

References

(References marked with an asterisk indicate studies that are included in this review.)

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of cross-cultural psychology*, *1*(3), 185-216.

Bufford, R. K., Paloutzian, R. F., & Ellison, C. W. (1991). Norms for the Spiritual Well Being Scale. *Journal of Psychology and Theology, 19*(1)*,* 56–70.

Cella, D., Lent, L., Bredle, J., Arnold, B., Pickard, L., & Romo, S. (2018). Overview of the FACIT-Sp. Retrieved from http://www.facit.org/FACITOrg/Overview

*Chimluang, J., Thanasilp, S., Akkayagorn, L., Upasen, R., Pudtong, N., & Tantitrakul, W. (2017). Effect of an intervention based on basic Buddhist principles on the spiritual well-being of patients with terminal cancer. *European Journal of Oncology Nursing*, *31*, 46-51. doi: 10.1016/j.ejon.2017.08.005

Chirico, F. (2016). Spiritual well-being in the 21st century: It's time to review the current WHO's health definition. *Journal of Health and Social Sciences*, *1*(1), 11-16. doi: 10.19204/2016/sprt2

*Chowdhury, R. M., & Fernando, M. (2013). The role of spiritual well-being and materialism in determining consumers' ethical beliefs: An empirical study with Australian consumers. *Journal of Business Ethics*, *113*(1), 61-79. doi:10.1007/s10551-012-1282x

Cobb, M., Puchalski, C., Rumbold, B. (2012). The future of spirituality and healthcare. In Cobb, M., Puchalski, C., Rumbold, B. (Eds.) *The Oxford Textbook of Spirituality in Health Care* (pp. 487-492). New York: Oxford University Press.

Cotton, S. P., Levine, E. G., Fitzpatrick, C. M., Dold, K. H., & Targ, E. (1999). Exploring the relationships among spiritual well-being, quality of life, and psychological adjustment in women with breast cancer. *Psycho - Oncology*, *8*(5), 429-438.

Ekşi, H., & Kardaş, S. (2017). Spiritual well-being: Scale development and validation. *Spiritual Psychology and Counseling*, *2*, 73-88. doi: 10.12738/spc.2017.1.0022

Ellison, C. W. (1983). Spiritual well-being: Conceptualization and measurement. *Journal of psychology and theology*, *11*(4), 330-338.

Ellison, C. W., & Smith, J. (1991). Toward an integrative measure of health and well-being. *Journal of Psychology and theology*, *19*(1), 35-45.

Fisher, J.W. (2011). The four domains model: Connecting spirituality, health and well being. *Religions*, *2*(1), 17-28. doi: 10.3390/rel2010017

Fisher, J. W., Francis, L. J., & Johnson, P. (2000). Assessing spiritual health via four domains of spiritual wellbeing: The SH4DI. *Pastoral psychology*, *49*(2), 133-145.

Genia, V. (2001). Evaluation of the spiritual well-being scale in a sample of college students. *The International Journal for the Psychology of Religion*, *11*(1), 25-33. doi:10.1207/515327582IJPR1101_03

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*(8). doi:10.1016/S0191-8869(03)00045-X

*Harrington, A. (2016). The importance of spiritual assessment when caring for older adults. *Ageing & Society*, *36*(1), 1-16. doi: 10.0144686X14001007

Hawley, J., & Hawley, J. A. (1993). *Reawakening the spirit in work: The power of dharmic management*. San Francisco, CA: Berrett-Koehler.

Heiser, D., & Brennan, M. (2013). Introduction: Spiritual Assessment and Intervention: Current Directions and Applications. In *Spiritual Assessment and Intervention with Older Adults* (pp. 9-28). New York, NY: Routledge.

McClain, C. S., Rosenfeld, B., & Breitbart, W. (2003). Effect of spiritual well-being on end of-life despair in terminally-ill cancer patients. *The lancet*, *361*. 1603-1607.

McDermott, M. A., & Palchanes, K. (1992). A process for translating and testing a quantitative measure for cross-cultural nursing research. *The Journal of the New York State Nurses' Association*, *23*(4), 12-15.

Moberg, D. O., & Brusek, P. M. (1978). Spiritual well-being: A neglected subject in quality of life research. *Social Indicators Research*, *5*(1-4), 303-323.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L.; Bouter, L. M., & De Vet, H. C. (2010a). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Quality of life research*, *19*(4), 539-549.

Mokkink, L. B., Terwee, C. B., Gibbons, E., Stratford, P. W., Alonso, J., Patrick, D. L., Knol, D. L., Bouter, L. M., & De Vet, H. C. (2010b). Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. *BMC Medical Research Methodology*, *10*(1), 82. doi: 10.1186/1471-2288-10-82

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2012). COSMIN checklist manual. *Amsterdam: University Medical Center*.

Monod, S., Brennan, M., Rochat, E., Martin, E., Rochat, S., & Büla, C. J. (2011). Instruments measuring spirituality in clinical research: a systematic review. *Journal of General Internal Medicine*, *26*(11), 1345. doi: 10.1007/s11606-011-1769-7

Morrison-Orton, D. J. (2004). How rehabilitation professionals define the concepts of spirituality and religion when working with individuals with disabilities. *Journal of Social Work in Disability & Rehabilitation, 3*(2), 37–55.

*Paloutzian, R., & Ellison, C. (1982). Loneliness, spiritual well-being and the quality of life In Peplau LA & Perlman D. (Eds.), *Loneliness: A sourcebook of current theory, research and therapy* (pp. 224–237). New York, NY: Wiley

*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 behavioral medicine*, *24*(1), 49-58.

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.

Riley, B. B., Perna, R., Tate, D. G., Forchheimer, M., Anderson, C., & Luera, G. (1998). Types of spiritual well-being among persons with chronic illness: Their relation to various forms of quality of life. *Archives of physical medicine and rehabilitation*, *79*(3), 258-264.

Serbetar, I., & Sedlar, I. (2016). Assessing reliability of a multi-dimensional scale by coefficient alpha. *Revija za Elementarno Izobrazevanje*, *9*, 189.

*Soleimani, M. A., Sharif, S. P., Allen, K. A., Yaghoobzadeh, A., Nia, H. S., & Gorgulu, O. (2017). Psychometric properties of the persian version of spiritual well-being scale in patients with acute myocardial infarction. *Journal of religion and health*, *56*(6), 1981-1997.

Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, *80*(1), 99-103.

Terwee, C. B., Mokkink, L. B., Knol, D. L., Ostelo, R. W., Bouter, L. M., & de Vet, H. C. (2012). Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Quality of Life Research*, *21*(4), 651-657.

Van Dierendonck, D. (2004). The construct validity of Ryff's Scales of Psychological Well being and its extension with spiritual well-being. *Personality and individual differences*, *36*(3), 629-643. doi: 10.1016/S0191-8869(03)00122-3

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: the kappa statistic. *Family Medicine*, *37*(5), 360-363.

Zinnbauer, B. J., Pargament, K. I., Cole, B., Rye, M. S., Butter, E. M., Belavich, T. G., Hipp, K. M., Scott, A. B. & Kadar, T. L. (1997). Religion and spirituality: Unfuzzying the fuzzy. *Journal for the Scientific Study of Religion, 36*(4), 549–564.

Appendix A – The extensive list of the frequencies of all extracted instruments

1	Functional Assessment of Chronic Illness Therapy— Spiritual Well-being Scale (FACIT-Sp)	153
2	Spiritual Well-Being Scale (SWBS) – Ellison and Paloutzian	129
3	Spiritual Well-Being Questionnaire (Fisher)	14
4	Multidimensional Measure of Religiousness and Spirituality (MMRS)	13
5	Spiritual health and life orientation measure (SHALOM, Fisher 1999)	12
6	Spiritual Wellness Inventory	2
7	Coping Strategies Questionnaire	2
8	Good Death Inventory	1
9	Intrinsic/Extrinsic Religiosity Scale	2
10	Index of Core Spiritual Experiences (INSPIRIT)	1
11	Religious Background and Behaviour Scale	2
12	Ways of religious coping scale	2
13	City of Hope-QOL Scale	3
14	Religious-Spiritual Well-Being	1
15	Quality of Life-Scale (QoL)	8
16	Resilience Scale (Young and Wagnild)	1
17	Attachement to God Inventory	1
18	Theodic Complaint Scale	1
19	Strength of the Satan Concept Scale	1
20	Spiritual Needs Questionnaire (SpRQ)	6
21	Spirituality/Religiousity and Coping (SpREUK-15)	2
22	Adaptive coping with Desease	1
23	Grace Scale (Richmont Grace Scale)	1
24	Spiritual care Inventory	2
25	Spiritual Care in Practice survey	1
26	Quality of Life Cancer Survivors Survey	6
27	Brief Religious Coping Scale	1
28	Religious Problem Solving Scale	1
29	Existential Sense of Meaning Scale	1
30	Physician’s Spiritual Well-Being Scale	2
31	Body-Mind-Spirit Well-Being Inventory (BMSWBI)	1

32	World Health Organization Quality of Life instrument (WHOQOL-SRPB)	6
33	"Spiritual Growth" subscale in the Health-Promoting Lifestyle Profile II (HPLP II)	1
34	Health-related quality of life (HRQOL)	4
35	Hope Scale (Willpower and Waypower Jerry Pattengale)	4
36	Spiritual Beliefs Inventory (SBI) (Holland et al.,1998)	1
37	Spiritual Involvement and Beliefs Scale (SIBS)	4
38	Spirituality Assessment Scale	1
39	Spiritual Health Inventory	1
40	Geriatric Spiritual Well-Being Scale	1
41	HBSC spirituality scale (adapted from Gomez)	1
42	Spiritual Perspective Scale	3
43	Feeling Good, Living Life (FGLL) (Fisher, 2004)	1
44	Spiritual Capital Questionnaire	1
45	Religious Involvement Inventory	1
46	Patient Dignity Inventory (PDI)	3
47	Mental, Physical and Spiritual Well-Being Scale	2
48	Religious commitment Inventory	1
49	Santa Clara Strength of Religious Faith Questionnaire	1
50	Springfield Religiosity Scale	1
51	Revised Religious Orientation Scale	1
52	Herth Hope Index	2
53	Multidimensional Inventory for Religious/Spiritual Well-Being	1
54	Sky Spirituality Scale (SS-25)	1
55	Religious Fundamentalism Scale	1
56	Survey of Health, Ageing, and Retirement in Europe (SHARE)	1
57	Brief Assessment of Spiritual Insight and Commitment Questionnaire	1
58	Children's Hope Scale	1
59	Spiritual Index of Well-being	4
60	Spiritual Distress Assessment Tool (SDAT)	1

Appendix B – FACIT-Sp Scale (12-item version)

This instrument is openly accessible in FACIT.org and may not be duplicated without permission. After registration on the website, the author of this review was allowed to use it.

FACIT-Sp-12 (Version 4)

Below is a list of statements that other people with your illness have said are important. Please circle or mark one number per line to indicate your response as it applies to the past 7 days.

		Not at all	A little bit	Some-what	Quite a bit	Very much
Sp1	I feel peaceful	0	1	2	3	4
Sp2	I have a reason for living	0	1	2	3	4
Sp3	My life has been productive	0	1	2	3	4
Sp4	I have trouble feeling peace of mind	0	1	2	3	4
Sp5	I feel a sense of purpose in my life	0	1	2	3	4
Sp6	I am able to reach down deep into myself for comfort	0	1	2	3	4
Sp7	I feel a sense of harmony within myself	0	1	2	3	4
Sp8	My life lacks meaning and purpose	0	1	2	3	4
Sp9	I find comfort in my faith or spiritual beliefs	0	1	2	3	4
Sp10	I find strength in my faith or spiritual beliefs	0	1	2	3	4
Sp11	My illness has strengthened my faith or spiritual beliefs	0	1	2	3	4
Sp12	I know that whatever happens with my illness, things will be okay	0	1	2	3	4

English (Universal)
Copyright 1987, 1997

Appendix C – SWBS (20 items)

This instrument was purchased from: https://www.lifeadvance.com/spiritual-well-being-scale.html and is digitally stamped with the buyers Name, Email and Transaction ID. It may not be replicated, copied or distributed beyond this review.

Buyer: Julia Pellengahr (julia.pellengahr@gmx.de)
Transaction ID: 8T8663658H7442144

SWB Scale

For each of the following statements circle the choice that best indicates the extent of your agreement or disagreement as it describes your personal experience:

SA	= Strongly Agree	D	= Disagree
MA	= Moderately Agree	MD	= Moderately Disagree
A	= Agree	SD	= Strongly Disagree

1.	I don't find much satisfaction in private prayer with God.	SA	MA	A	D	MD	SD
2.	I don't know who I am, where I came from, or where I'm going.	SA	MA	A	D	MD	SD
3.	I believe that God loves me and cares about me.	SA	MA	A	D	MD	SD
4.	I feel that life is a positive experience.	SA	MA	A	D	MD	SD
5.	I believe that God is impersonal and not interested in my daily situations.	SA	MA	A	D	MD	SD
6.	I feel unsettled about my future.	SA	MA	A	D	MD	SD
7.	I have a personally meaningful relationship with God.	SA	MA	A	D	MD	SD
8.	I feel very fulfilled and satisfied with life.	SA	MA	A	D	MD	SD
9.	I don't get much personal strength and support from my God	SA	MA	A	D	MD	SD
10.	I feel a sense of well-being about the direction my life is headed in.	SA	MA	A	D	MD	SD
11.	I believe that God is concerned about my problems.	SA	MA	A	D	MD	SD
12.	I don't enjoy much about life.	SA	MA	A	D	MD	SD
13.	I don't have a personally satisfying relationship with God.	SA	MA	A	D	MD	SD
14.	I feel good about my future.	SA	MA	A	D	MD	SD
15.	My relationship with God helps me not to feel lonely.	SA	MA	A	D	MD	SD
16.	I feel that life is full of conflict and unhappiness.	SA	MA	A	D	MD	SD
17.	I feel most fulfilled when I'm in close communion with God.	SA	MA	A	D	MD	SD
18.	Life doesn't have much meaning.	SA	MA	A	D	MD	SD
19.	My relation with God contributes to my sense of well-being.	SA	MA	A	D	MD	SD
20.	I believe there is some real purpose for my life.	SA	MA	A	D	MD	SD

SWB Scale © 1982 by Craig W. Ellison and Raymond F. Paloutzian. All rights reserved. Not to be duplicated unless express written permission is granted by the authors or by Life Advance. See www.lifeadvance.com.

Appendix D – Box A, B and D of the COSMIN checklist

Step 3. Determining if a study meets the standards for good methodological quality

Box A. Internal consistency				
	excellent	**good**	**fair**	**poor**
1 Does the scale consist of effect indicators, i.e. is it based on a reflective model?				
Design requirements				
2 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
3 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
4 Was the sample size included in the internal consistency analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
5 Was the unidimensionality of the scale checked? i.e. was factor analysis or IRT model applied?	Factor analysis performed in the study population	Authors refer to another study in which factor analysis was performed in a similar study population	Authors refer to another study in which factor analysis was performed, but not in a similar study population	Factor analysis NOT performed and no reference to another study
6 Was the sample size included in the unidimensionality analysis adequate?	7* #items and ≥100	5* #items and ≥100 OR 6-7* #items but <100	5* #items but <100	<5* #items

7	Was an internal consistency statistic calculated for each (unidimensional) (sub)scale separately?	Internal consistency statistic calculated for each subscale separately			Internal consistency statistic NOT calculated for each subscale separately
8	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
9	for Classical Test Theory (CTT), continuous scores: Was Cronbach's alpha calculated?	Cronbach's alpha calculated		Only item-total correlations calculated	No Cronbach's alpha and no item-total correlations calculated
10	for CTT, dichotomous scores: Was Cronbach's alpha or KR-20 calculated?	Cronbach's alpha or KR-20 calculated		Only item-total correlations calculated	No Cronbach's alpha or KR-20 and no item-total correlations calculated
11	for IRT: Was a goodness of fit statistic at a global level calculated? E.g. χ^2, reliability coefficient of estimated latent trait value (index of (subject or item) separation)	Goodness of fit statistic at a global level calculated			Goodness of fit statistic at a global level NOT calculated

NB. Item 1 is used to determine whether internal consistency is relevant for the instrument under study. It is not used to rate the quality of the study.

Box B. Reliability: relative measures (including test-retest reliability, inter-rater reliability and intra-rater reliability)				
	excellent	**good**	**fair**	**poor**
Design requirements				
1 Was the percentage of missing items given?	Percentage of missing items described	Percentage of missing items NOT described		
2 Was there a description of how missing items were handled?	Described how missing items were handled	Not described but it can be deduced how missing items were handled	Not clear how missing items were handled	
3 Was the sample size included in the analysis adequate?	Adequate sample size (≥100)	Good sample size (50-99)	Moderate sample size (30-49)	Small sample size (<30)
4 Were at least two measurements available?	At least two measurements			Only one measurement
5 Were the administrations independent?	Independent measurements	Assumable that the measurements were independent	Doubtful whether the measurements were independent	measurements NOT independent
6 Was the time interval stated?	Time interval stated		Time interval NOT stated	
7 Were patients stable in the interim period on the construct to be measured?	Patients were stable (evidence provided)	Assumable that patients were stable	Unclear if patients were stable	Patients were NOT stable
8 Was the time interval appropriate?	Time interval appropriate		Doubtful whether time interval was appropriate	Time interval NOT appropriate
9 Were the test conditions similar for both measurements? e.g. type of administration, environment, instructions	Test conditions were similar (evidence provided)	Assumable that test conditions were similar	Unclear if test conditions were similar	Test conditions were NOT similar

10	Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study
Statistical methods					
11	for continuous scores: Was an intraclass correlation coefficient (ICC) calculated?	ICC calculated and model or formula of the ICC is described	ICC calculated but model or formula of the ICC not described or not optimal. Pearson or Spearman correlation coefficient calculated with evidence provided that no systematic change has occurred	Pearson or Spearman correlation coefficient calculated WITHOUT evidence provided that no systematic change has occurred or WITH evidence that systematic change has occurred	No ICC or Pearson or Spearman correlations calculated
12	for dichotomous/nominal/ordinal scores: Was kappa calculated?	Kappa calculated			Only percentage agreement calculated
13	for ordinal scores: Was a weighted kappa calculated?	Weighted Kappa calculated		Unweighted Kappa calculated	Only percentage agreement calculated
14	for ordinal scores: Was the weighting scheme described? e.g. linear, quadratic	Weighting scheme described	Weighting scheme NOT described		

Box D. Content validity (including face validity)				
	excellent	**good**	**fair**	**poor**
General requirements				
1 Was there an assessment of whether all items refer to relevant aspects of the construct to be measured?	Assessed if all items refer to relevant aspects of the construct to be measured		Aspects of the construct to be measured poorly described AND this was not taken into consideration	NOT assessed if all items refer to relevant aspects of the construct to be measured
2 Was there an assessment of whether all items are relevant for the study population? (e.g. age, gender, disease characteristics, country, setting)	Assessed if all items are relevant for the study population in adequate sample size (≥10)	Assessed if all items are relevant for the study population in moderate sample size (5-9)	Assessed if all items are relevant for the study population in small sample size (<5)	NOT assessed if all items are relevant for the study population OR target population not involved
3 Was there an assessment of whether all items are relevant for the purpose of the measurement instrument? (discriminative, evaluative, and/or predictive)	Assessed if all items are relevant for the purpose of the application	Purpose of the instrument was not described but assumed	NOT assessed if all items are relevant for the purpose of the application	
4 Was there an assessment of whether all items together comprehensively reflect the construct to be measured?	Assessed if all items together comprehensively reflect the construct to be measured		No theoretical foundation of the construct and this was not taken into consideration	NOT assessed if all items together comprehen-sively reflect the construct to be measured
5 Were there any important flaws in the design or methods of the study?	No other important methodological flaws in the design or execution of the study		Other minor methodological flaws in the design or execution of the study	Other important methodological flaws in the design or execution of the study