

This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Instruments to evaluate mental well-being in old age: a systematic review

Martín-María, Natalia; Lara, Elvira ; CresswellSmith, Johanna; Forsman, Anna; Kalseth, Jorid; Valeria, Donisi; Francesco, Amaddeo; Wahlbeck, Kristian; Miret, Marta

Published in:
Aging and Mental Health

DOI:
[10.1080/13607863.2020.1774742](https://doi.org/10.1080/13607863.2020.1774742)

Published: 01/01/2020

Document Version
Accepted author manuscript

Document License
Unknown

[Link to publication](#)

Please cite the original version:

Martín-María, N., Lara, E., CresswellSmith, J., Forsman, A., Kalseth, J., Valeria, D., Francesco, A., Wahlbeck, K., & Miret, M. (2020). Instruments to evaluate mental well-being in old age: a systematic review. *Aging and Mental Health*. <https://doi.org/10.1080/13607863.2020.1774742>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Title

Instruments to evaluate mental well-being in old age: a systematic review

Authors

Natalia Martín-María^{1,2,3}; Elvira Lara^{1,2,3}; Johanna Cresswell-Smith⁴; Anna K. Forsman⁵; Jorid Kalseth⁶; Valeria Donisi⁷; Francesco Amaddeo⁷; Kristian Wahlbeck⁴; Marta Miret^{1,2,3*}

Affiliations

¹ Department of Psychiatry, Universidad Autónoma de Madrid, Spain.

² Instituto de Salud Carlos III, Centro de Investigación Biomédica en Red de Salud Mental. CIBERSAM, Madrid, Spain.

³ Department of Psychiatry, Hospital Universitario de La Princesa, Instituto de Investigación Sanitaria Princesa (IIS-Princesa), Madrid, Spain.

⁴ Mental Health Unit, Finnish Institute for Health and Welfare (THL), Helsinki, Finland.

⁵ Faculty of Education and Welfare Studies, Health Sciences, Åbo Akademi University, Vaasa, Finland.

⁶ Department of Health Research, SINTEF Technology and Society, Trondheim, Norway.

⁷ Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Italy.

*Corresponding author

Marta Miret, PhD, Department of Psychiatry, Universidad Autónoma de Madrid, Arzobispo Morcillo Street 4, 28029 Madrid, Spain. E-mail: marta.miret@uam.es

Funding. This study was funded by The Joint Programming Initiative (JPI) “*More Years, Better Lives- The Potential and Challenges of Demographic Change*”; and has received funding by the “*Acciones de Programación Conjunta Internacional 2016*” program (PCIN-2016-118) of the Spanish Research Agency (AEI) of the Spanish Ministry of Sciences and Innovation, and by the Instituto de Salud Carlos III-FIS (research grant PI16/00177); the Italian Ministry of Health; the Academy of Finland; and the Research Council of Norway; NMM is supported by the programme ‘Contratos predoctorales para Formación de Personal Investigador, FPI-UAM’, Universidad Autónoma de Madrid, Spain. EL’s work is supported by the Sara Borrell postdoctoral

programme (CD18/00099) from the Instituto de Salud Carlos III (Spain) and co-funded by European Union (ERDF/ESF, “Investing in your future”).

Declaration of interest statement. None.

Total word count (excluding appendices, references, tables and figures): 4952

Abstract

Objective. The aim of this study is to identify and appraise existing instruments to evaluate mental well-being in old age.

Method. Systematic literature searches in PubMed, PsycINFO, ProQuest Research Library, AgeLine, and CINAHL databases were performed. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline was used to assess the measurement properties, reported according to the Preferred Reporting Items for Systematic Reviews and meta-Analysis (PRISMA) statement. For each measurement property, results were classified as positive, negative, or indeterminate. The quality level of evidence was rated as high, moderate, low, or very low following the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach.

Results. A total of 28 instruments were found. Most instruments evaluated different dimensions of mental well-being, including various subscales. The quality was adequate overall. Six instruments showed high quality (Perceived Well-Being Scale-PWB, Salamon-Conte Life Satisfaction in the Elderly Scale-SCLSES, Herth Hope Scale-HHS, Life Satisfaction Index Third Age-LSITA, Meaning in Life Scale-MLS, and SODdisfazione dell'Anziano-SODA), and other six a moderate level (Scale of Happiness of the Memorial University of Newfoundland-MUNSH, Six Scales of Psychological Well-Being-PWBS, Valuation Of Life-VOL, Life Satisfaction Scale for Chinese Elders-LSS-C, Meaningful Activity Participation Assessment-MAPA, and Will To Life-WTL).

Conclusion. This review provides the first comprehensive synthesis of instruments assessing mental well-being in older populations. The PWB, SCLSES, HHS, LSITA, MLS and SODA were the most appropriated instruments. An instrument that specifically measures mental well-being in the oldest old age group (aged 80 plus) and that considers its multidimensional nature is needed.

Keywords: older adults; measurement tools; evaluative well-being; hedonic well-being; eudaimonic well-being.

Introduction

Mental well-being is a core component of general well-being at all ages. In old age, mental well-being is strongly associated with a higher level of functioning and better health outcomes (Fox, Stathi, McKenna, & Davis, 2007). It also has a bidirectional connection with physical well-being, and is linked to prolonged life expectancy (Steptoe, Demakakos, & de Oliveira, 2012).

Due to the ongoing demographic transition, the population of older adults is growing. In 2018, the global population aged 65 and over outnumbered children under five years and is projected to more than double by 2050 when it is expected to reach more the 1.5 billion (United Nations, 2019). This demographic shift brings with it new challenges and the need for increased understanding of the ageing process. In order to develop the knowledge of mental well-being in older age, valid and reliable measures specifically for the older population are needed. Measuring mental well-being in old age may not only benefit health outcomes, reduce disability and increase autonomy, but also provide important information for public health endeavours, such as service provision and shaping policy (Steptoe, Deaton, & Stone, 2015).

Mental well-being in older age has been found to be both dynamic and multi-dimensional (Lara et al., 2019) along three perspectives: evaluative well-being (or life satisfaction) (Diener, Emmons, Larsen, & Griffin, 1985), hedonic well-being (the moment-to-moment experience of positive emotions) (McMahan & Estes, 2012), and eudaimonic well-being (competence and personal development, meaning in life, and fulfilment of the own potential) (Ryan & Deci, 2001; Ryff, 1989).

Previous research has found that older adults tend to experience greater positive emotions (Lawton, Kleban, Rajagopal, & Dean, 1992) despite lower levels of life satisfaction than their younger counterparts (Deaton, 2008). This has been related to a stronger focus on the positive aspects of life in older age, with less focus on the negative aspects (Mather & Carstensen, 2005). One possible explanation for this phenomenon is the socioemotional selectivity theory (Carstensen, Fung, & Charles, 2003), which states that older adults appreciate their time as limited and therefore prioritise short time goals in relation to social connections, and regulate emotional states to optimise mental well-being. Older adults therefore shift their motivational priorities, and place greater emphasis on deepening existing relationships and developing expertise in already satisfying areas of life. Furthermore, experiencing a sense of purpose in life and

continuing personal growth may contribute to improved health outcomes also in older age (Zaslavsky et al., 2014).

Mental well-being measures such as the Life Satisfaction Index (LSI) (Neugarten, Havighurst, & Tobin, 1961) and the Life Orientation Test (LOT) (Scheier, Carver, & Bridges, 1994) are often devised to evaluate the general population. Specific instruments or scales assessing mental well-being in older age appear to be lacking, although these are needed considering the circumstances which often surround older age such as age specific demands and losses (Cresswell-Smith et al., 2019; Lara et al., 2019; Poon & Cohen-Mansfield, 2011).

Previous reviews have examined different measures of well-being in older age including quality of life (Makai, Brouwer, Koopmanschap, Stolk, & Nieboer, 2014), and health and well-being tools for public health and community interventions in the general population (Cooke, Melchert, & Connor, 2016; Dronavalli & Thompson, 2015; Kobau, Sniezek, Zack, Lucas, & Burns, 2010; McDowell, 2010). A review looking at measurement and prediction of mental well-being in later life (McNeil, Stones, & Kozma, 1986) has also been performed, although this was published more than 30 years ago. The present review is therefore timely both in terms of identifying which instruments may be most suited for the older population, but also in order to guide public policy development towards an increased focus on mental well-being in the older population.

The aim of the present study is to systematically review and evaluate existing instruments for measuring mental well-being in old age.

Method

This review followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) recommendations (Liberati et al., 2009; Moher, Liberati, Tetzlaff, & Altman, 2009). The protocol of this systematic review has been previously registered at the International Prospective Register of Systematic Reviews (PROSPERO) (registration number: CRD42018110653).

Data sources and searches

A literature search using a combination of various keywords of instruments and mental well-being in the ageing population was carried out. The specific search strategy for each database is described in the Appendix. The following electronic databases were used: (1) PubMed, by the National Library of Medicine (NLM); (2) PsycINFO, by the

American Psychological Association (APA PsyNET); (3) ProQuest Research Library, by Cambridge Information Group; (4) AgeLine, by EBSCO; and (5) CINAHL, by EBSCOhost. Additionally, a ‘snow-balling’ technique was combined with the manual search to locate primary instruments which had not been identified in the bibliographic database searches: e.g., articles which used some of these instruments to test the convergent validity of other scale, books addressing the topic of instruments to measure mental well-being, or reviews focusing on the description of instruments to evaluate mental well-being throughout the life-course.

Criteria for considering instruments for the review

The inclusion criteria for the instruments were the following:

Study population: ageing people (no lower limit).

Study requirements: (1) psychometric studies on the development of instruments evaluating mental well-being; (2) published in peer-reviewed journals; (3) studies that provide at least the reliability or the validity of those instruments; (4) no language or time restrictions were applied. Published instruments until April 2018 were included.

The exclusion criteria for the instruments were the following:

Study requirements: (1) psychometric studies on the development of instruments measuring negative affect, religiosity, quality of life, self-rated health, successful ageing, functional status, self-esteem, resilience, locus of control and satisfaction with specific domains (i.e. job, couple, or care); (2) subsequent validations and applications of the original instruments.

Data collection procedure

A three-step process was followed by three researchers. Firstly, N.M.-M. retrieved publications which evidenced instruments evaluating mental well-being in old age and transferred them to EndNote reference management software (Thomson Reuters, 2013) in order to remove the duplicates.

In a second step, articles were identified by reading the title and abstract. Articles that did not meet the inclusion criteria were excluded. A random sample comprising 20% of the total abstracts were double-checked by a second researcher (E.L.). Initial disagreements were resolved via discussion between the two reviewers; if necessary, a third investigator was consulted (M.M.).

In a third step, the full text of the included articles were read and the following information extracted: name of the instrument, authors, publication date, and details of

the study sample (number of participants, age, gender, nation, and whether the sample came from the general population or it was other specific population). Instruments were also classified by the aspect of mental well-being assessed: evaluative (life satisfaction), hedonic (positive emotions), eudaimonic (personal development, meaning in life, and achievement of one's potential), or a combination of these aspects, labelled as 'multi-dimensional' classification. The latter category covered instruments with no subscales that mixed items of different aspects of mental well-being, and instruments with distinct subscales incorporating questions of different aspects within the same subscale. Finally, the interpretability and feasibility (type and ease of application, cost, length, and completion time) of each instrument were also assessed, because although they are not measurement properties, they are important aspects to differentiate between instruments of similar quality (Prinsen et al., 2016).

Psychometric properties and quality appraisal

The psychometric properties and quality of each instrument was then critically appraised following an adapted version of the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guideline for systematic reviews of patient-reported outcome measures (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), which has obtained international acceptance and follows a systematic methodology (Mokkink et al., 2010). The COSMIN tool covers nine different domains: content validity, structural validity, hypothesis testing for construct validity, criterion validity, internal consistency, test-retest reliability, measurement error, cross-cultural validity, and responsiveness. Content validity assesses whether the instrument is an adequate reflection of the construct being measured; it is considered the most important measurement property, so that a lack of it can affect the remaining measurement properties. Structural validity is defined as the extent to which the scores of an instrument are an adequate reflection of the dimensionality of the construct being assessed. Hypothesis testing for construct validity on the other hand (including both convergent and known groups validity), relates to whether the scores of an instrument are consistent with hypotheses, based on the assumption that the instrument validly measures the construct to be measured. Criterion validity is the degree to which the scores of a scale adequately reflect a 'gold standard' or a longer version of the scale. Internal consistency is defined as the degree to which the instrument scales correlate, whereas test-retest reliability concerns the degree to which scores for individuals remain the same in repeated measurements under different conditions. Measurement error is

defined as the random and systematic difference between a participant's score and its true value. Cross-cultural validity is the degree to which the performance of the items on an adapted instrument are an adequate reflection of the original version. Finally, responsiveness is defined as the ability of an instrument to detect clinical change over time (Mokkink, Prinsen, Bouter, de Vet, & Terwee, 2016). Each instrument was individually coded using a positive '+' code to denote an adequate description, value, measure or argument in relation to the psychometric property, and a negative code '-' to denote inadequacy or what was considered to be below acceptable standards in each psychometric property. Finally, an indeterminate code '?' was used to denote uncertain methods, measures or design, or if no information was available about a particular psychometric property. When articles reported the internal consistency for different subscales of one instrument, the lowest score was considered. In the same way, when more than one gold standard was used for testing criterion validity, if either of them reached a correlation ≥ 0.70 , this score was considered. Finally, when the instruments informed about correlations with subscales as well as with the total scale, the latter score was selected. Detailed COSMIN criteria for measurement properties can be found in Table S1. None of the reviewed instruments reported data on measurement error, cross-cultural validity or responsiveness; therefore, those properties were not included in the current review. Nevertheless, available translations of each instrument were provided. Finally, a modified Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach from the COSMIN methodology (Guyatt et al., 2008; Prinsen et al., 2018) was used to rate the quality of the evidence available for the measurement properties of each instrument. This approach takes into account the methodological quality, the consistency of results, and the directness and precision of evidence. The overall quality of evidence was rated as high, moderate, low, or very low. The derived information was analysed and a comparison between instruments was carried out based on all the above-mentioned criteria and characteristics.

Results

A total of 3684 articles were found based on the initial search strategy (Figure 1). After excluding duplicates, 2250 articles were screened. Of these, 229 articles were retrieved for more detailed evaluation according to title and abstract. A total of 127 articles were excluded for different reasons, including being out of scope, not having an aging sample, or not measuring mental wellbeing. Validations in other languages or

applications in specific populations of the original instruments (82 articles), which were not the focus of the present review, were also excluded. Finally, 20 original instruments designed for older population were found. Eight instruments originally created for older adults were identified in the manual search (see Table 1).

PLEASE, INSERT FIGURE 1 APROXIMATELY HERE

The degree of agreement between the researchers was 97.4%, which indicates a high level of inter-rater agreement (Peat, 2001). The Kappa coefficient was 0.89 (95% CI =0.82, 0.95).

PLEASE, INSERT TABLE 1 APROXIMATELY HERE

A total of 13 of the selected instruments were created prior to the year 2000, with the remaining 15 after that. Instruments measuring the eudaimonic well-being aspect (42.8%), and multi-dimensional instruments (39.3%) were the most common. In the 11 multi-dimensional instruments, the eudaimonic well-being aspect was always included, the evaluative aspect was included in ten instruments, and the hedonic category only in 8 instruments. Those exclusively measured the evaluative well-being aspect were less common, representing 14.3% of the total. Only one instrument measured hedonic well-being exclusively (3.7%), which was the Con-Dis device for measuring perceived well-being (Reijula et al., 2009).

The included instruments were applied to a wide age range, from 50 to 78 years old. The most frequently assessed age-group was the 60-69 group (57.1%), followed by the 50-59 years age group (17.9%). Studies on five of the instruments (17.9%), did not specify the age range of participants, only mentioning a focus on the older population. Only the VOL (Lawton et al., 2001) and WTL (Carmel, 2017) instruments were created specifically for people aged 70+ and 78+, respectively.

All instruments were validated in both genders, (except for the Life-satisfaction construct (Closs & Kempe, 1986), which was assessed only for women) and in general and institutionalised population except for two instruments, which measured spiritual well-being (SIWB and TSWBATECI) in chronically ill participants. Four instruments (SWLS, PWBS, HHS, and the Con-Dis device), were validated in less than 100 participants. The smallest sample comprised ten individuals (Con-Dis device), whereas the largest sample comprised 2,160 respondents (TSWBATECI).

A total of 16 instruments originated from the United States of America, other six were created in Asia, and the remaining six in Europe. English was the original language for most of the instruments (89.2%), although subsequent translations are

available in different languages. SWLS (Diener et al., 1985) was the most translated instrument, with more than 30 distinct versions used globally, followed by PSWB (Ryff, 1989), HHS (Herth, 1991), SHS (Lyubomirsky & Lepper, 1999), and CASP (Hyde, Wiggins, Higgs, & Blane, 2003). Most instruments include subscales, except for the Life-Satisfaction construct (Closs & Kempe, 1986), the Congruity Life Satisfaction Scale (Meadow, Mentzer, Rahtz, & Sirgy, 1992), SHS (Lyubomirsky & Lepper, 1999), and MAPA (Eakman, Carlson, & Clark, 2010), and most of the response scales were Likert type (only 25.0% of them included categorical answers). Regarding the feasibility aspects, the majority of instruments were self-administered (except for the Life-Satisfaction construct (Closs & Kempe, 1986) and the Integration Inventory (Ruffing-Rahal, 1991), that were interviewer-administered), and were available in the public domain, excluding SCLSES (Salamon, 1988), CAWP (Ku, Fox, & McKenna, 2008), LSS-C (Lou, Chi, & Mjelde-Mossey, 2008) and LSITA, which was available upon request to the authors by email (Barrett & Murk, 2006). The number of items per instrument ranged from four (SHS) to 200 (SCLSES).

Quality appraisal

The quality appraisal of the instruments is shown in Table 2, whereas their psychometric properties is specified in Table S2.

PLEASE, INSERT TABLE 2 APROXIMATELY HERE

Content validity

In almost all cases (92.9%), the articles gave a clear description of the relevance (whether relevant items were included for the construct, for the target population and for the context of interest, and whether there are appropriate response options and an appropriate recall period), comprehensiveness (if all key concepts were included) and comprehensibility (whether instructions, items and response options are understandable by the target population, the items are appropriately worded, and the options match the question as intended). However, a relevant selection of items was not included in the design of the instruments in all instances (RPGCMS (Lawton, 1975) and SIWB (Daaleman, Frey, Wallace, & Studenski, 2002)).

Structural validity

SCLSES (Salamon, 1988), Life-satisfaction construct, PWBS (Ryff, 1989), HHS (Herth, 1991), VOL (Lawton et al., 2001), the Thai Elders Psychological Well-Being measure (Ingersoll-Dayton, Saengtienchai, Kespichayawattana, & Aunguroch, 2004), LSITA (Barret & Murk, 2006), CAWP (Ku et al., 2008), LSS-C (Lou et al., 2008),

CASP-19 (Hyde et al., 2003), TSWBATECI (Unsanit, Sunsem, Kunsongkeit, O'Brien, & McMullen, 2012), WTL (Carmel, 2017), MLS (Lee & Hong, 2017), , and SODA (Fastame, Penna, & Hitchcott, 2019) obtained a positive value for this property, with factors explaining more than 50% of the variance, CFI (comparative fit index) or TLI (Tucker-Lewis index) > 0.95, or RMSEA (root mean square error of approximation) or SRMR (standardized root mean residuals) < 0.08. The remaining instruments presented lower structural validity values, whereas several instruments (35.7%) did not provide such information.

Hypothesis testing for construct validity

Results were in accordance with the presented hypothesis in 39.3% of cases, although the remaining instruments did not report the existence of a hypothesis.

Criterion validity

Criterion validity was evaluated by Spearman's correlation with different gold standard instruments. Almost one third of the instruments (32.1%) obtained a positive rating (HHS, SHS, MLS, Con-Dis device, RPCGMS, MUNSH, SCLSES, PWB and LSITA), with values greater than 0.70. More than half (57.1%), presented values under 0.70, and the remaining instruments did not provide data in this area.

Internal consistency

Internal consistency was evaluated by Cronbach's alpha, either for the entire instrument or for each subscale. Almost all the instruments (82.1%) obtained a positive rating, with values ranging from 0.70 to 0.98. Four instruments (SWLS, Life-satisfaction construct, CASP-19, and SCLSES) had a negative rating, because they had some subscales with values below 0.70. No information was available for the Con-Dis device (Reijula et al., 2009).

Test-retest reliability

SWLS (Diener et al., 1985), LSS-C (Lou et al., 2008), PWBS (Ryff, 1989), HHS (Herth, 1991), LSS-C (Lou et al., 2008), MAPA (Eakman et al., 2010), MLS (Lee & Hong, 2017), , and SODA (Fastame et al., 2019) obtained a positive value for this property, with an intraclass correlation coefficient (ICC) or a Pearson's correlation greater than 0.70 and 0.80, respectively. Nevertheless, RPGCMS (Lawton, 1975), SCLSES (Salamon, 1988) MUNSH (Kozma & Stones, 1980) and PWB (Reker & Wong, 1984) correlations ($r = 0.75, 0.67, 0.70$ and 0.78 respectively) were also rated with a positive value due to their good rates over a long period of time (three months,

six months, one year, and up to two years in the last case). The remaining instruments had lower values, whereas 42.9% did not provide information in this respect.

Quality level of evidence

The quality level of evidence of the included instruments was adequate, with six showing high quality (HHS, MLS, SCLSES, PWB, LSITA, and SODA), and other six reaching a moderate level (LSS-C, PWBS, VOL, MAPA, WTL and MUNSH) (Table2).

The LSS-C (Lou et al., 2008), measuring evaluative well-being, stands out as the best of its category, although with a moderate level of quality. It was validated in a large sample of individuals, does not take long to administer (8 items), was culturally sensitive, and recommended for use on Chinese older adults. However, is not available in the public domain (Lou et al., 2008). In terms of eudaimonic well-being, the HHS (Herth, 1991) as well as the MLS (Lee & Hong, 2017) were the most highly rated instruments, followed by PWBS (Ryff, 1989), VOL (Lawton et al., 2001), MAPA (Eakman et al., 2010), and WTL (Carmel, 2017), that obtained a moderate level of quality. The HHS (Herth, 1991) measured hope using 30 items, was available in the public domain, and reached excellent quality rates in all the psychometric domains, although the sample size was small and hypothesis were not defined. On the other hand, the MLS (Lee & Hong, 2017), which measures meaning in life with 12 items, achieved the best possible positive ratings, although it was only available in Korean. The only instrument assessing hedonic well-being was the Con-Dis device (Reijula, 2009), which was validated in a reduced sample and only obtained positive values in content and in criterion validity, resulting in a low level of quality. Within the multi-dimensional category, four instruments achieved the highest level of quality: SCLSES (Salamon, 1988), PWB (Reker & Wong, 1984), LSITA (Barret & Murk, 2006), and SODA (Fastame et al., 2019), followed by MUNSH (Kozma & Stones, 1980), that obtained a moderate quality. The SCLSES (Salamon, 1988) was a 200 question instrument designed to assess life satisfaction, so it requires a long administration time, and furthermore, it was not available in the public domain. The PWB (Reker & Wong, 1984) reached a positive rating in all domains except for the structural validity (reaching almost 50% of explained variance), and it allowed assessments of psychological and physical well-being separately. The LSITA (Barret & Murk, 2006) measured life satisfaction, was available to researchers upon request to the authors by email, and reached good validity and internal consistency, although test-retest reliability was not reported. Finally, the SODA (Fastame et al., 2019), with 14 items, is a brief, reliable,

and valid instrument for the assessment of satisfaction, although it was only available in Italian.

Discussion

The current synthesis of literature shows some reliable and valid instruments for evaluating mental well-being in older populations. The results suggest significant improvements in the development of mental well-being measurement instruments in recent decades since the first one published in 1975 (Lawton). The majority of the instruments have been validated in healthy and wealthy populations from North America, Asia and Europe, and none of them originated from Africa, Australia or South America, making their applicability in other contexts debatable (Torres, 1999).

An important aspect to consider when assessing these instruments in the context of older age is the different age ranges varying from 50 to 78+ years. This scenario is a typical reflection of the general disagreement and the difficulties in defining old age. Most high and middle-income nations assume the chronological age of 65 years (traditionally corresponding with retirement), as a definition of old age. This classification does not adapt well to the situation in many low-income nations (Kowal & Dowd, 2001), where old age may begin when active contribution is no longer possible and new roles emerge, that may correspond to 50 years and older (Gorman, 1999).

Interestingly, none of the mental well-being instruments were specifically created for the oldest-old population which is often defined as 80 years and older, a segment of the population which is projected to increase dramatically in future decades (He, Goodkind, & Kowal, 2016). This also reflects the general lack of research in the oldest age group (Liljas et al., 2017). Therefore, reliable and valid instruments measuring mental well-being with a sample specifically of older adults aged 80+ are needed. These knowledge gaps could benefit from further research especially taking into account that this population is particularly vulnerable to mental and physical illnesses (Cohen-Mansfield et al., 2013). Developing such measures could contribute towards increased attention to and better chances of preserving and improving mental well-being in the oldest-old age (Steptoe et al., 2015). Future studies aiming to design an instrument for the oldest old population should consider feasibility aspects (self- or interviewer- administered, available in the public domain, and short in terms of number of items and administration time) and take into account all the mental well-being dimensions -functional, social, personal and environmental-, that are relevant for individuals aged 80+ (Lara et al., 2019).

The dynamic and multi-dimensional nature of mental well-being (Lara et al., 2019) is reflected in both the instruments evaluating different aspects of mental well-being simultaneously and the existence of subscales in most of the reviewed instruments. Subscales measuring mood, perceived health, social relationships, and values of life were the most common within the instruments. In spite of the multi-dimensional nature of mental well-being there is only one instrument specifically assessing the hedonic well-being aspect. This may be due to the fact that momentary assessments of mood may not require age specific questions, and can be evaluated in any age group. However, it could be argued that more daily and momentary evaluations throughout the entire day are needed to better understand the hedonic experiences, especially in the older population (Heo, Kim, Kim, & Heo, 2014). Testing the multi-dimensional model through structural equation modelling might help identify constructs as inputs, mediators, moderators or outcomes of mental well-being (Jayawickreme, Forgeard, & Seligman, 2012). Future research should attempt to overcome the demographic limitation by developing instruments in cultural contexts where they do not yet exist, assessing their psychometric properties as well. Equally important is to include the target group in the development process, this is to say, create instruments for and with the older adults to inform about their needs.

Given the objective of the present review, to identify and appraise existing instruments to evaluate mental well-being in old age, the most suitable instruments are presented by the aspect of mental well-being assessed, according to the purpose, quality of the measurement properties, and feasibility aspects. Regarding evaluative and experienced well-being, no instrument is recommended, since none of them reached a high level of quality. Nevertheless, two instruments obtained good quality rates in terms of eudaimonic well-being: the HHS (Herth, 1991) and the MLS (Lee & Hong, 2017). The HHS (Herth, 1991) might be the instrument of choice. Although hypotheses were not defined, it is available in several languages, whereas MLS is only available in Korean. Finally, four instruments achieved the highest level of quality within the multi-dimensional category: SCLSES (Salamon, 1988), PWB (Reker & Wong, 1984), LSITA (Barret & Murk, 2006), and SODA (Fastame et al., 2019). Given that the SODA (Fastame et al., 2019) was only available in Italian and that the SCLSES (Salamon, 1988) comprised 200 items, the PWB and the LSITA might be the instruments of choice. The LSITA (Barret & Murk, 2006), with 35 items and validated in a sample of individuals aged 50+, could be used with middle-age older people, cognitively healthy,

whereas the PWB (Reker & Wong, 1984), with 14 items and validated in a sample of individuals aged 62+, could be employed with healthy older population or with a minimal cognitive decline.

Nonetheless, it should be noted that none of the reviewed instruments reported data on measurement error, responsiveness or cross-cultural validity in their original version (e.g., SWLS has been translated and validated in several languages afterwards). Since longitudinal studies to measure change over time and subsequent validations and applications of the original instruments were not included in the present review, cross-cultural validity and responsiveness were not assessed in any of the instruments. However, measurement error should have been analysed, since its presence could threaten the validity of the instrument and the interpretability of the scores.

Previous reviews of instruments were focused only on adolescents (Rose et al., 2017; Tsang, Wong, & Lo, 2012), whereas a recent systematic review of well-being measurement scales in the general and healthy population (Lindert, Bain, Kubzansky, & Stein, 2015), was not able to recommend any particular scale, due to the limited information available for the instruments identified. To the best of our knowledge, this is the first systematic review of instruments evaluating mental well-being in the older population. The present study has several strengths, including the use of the most up-to-date version of COSMIN guidelines for conducting systematic reviews (Prinsen et al., 2018) and the GRADE approach used to rate the quality of the evidence (Terwee et al., 2018), the large number of instruments analysed, as well as the focus specifically on the older population. Additionally, this review includes publications written in any language, offering a more comprehensive and global picture of the existing instruments for measuring mental well-being at old age. A good level of agreement was found between researchers conducting the data extraction. Moreover, articles were included irrespective of the sample size in order to cover all available instruments; although it should be noted that small sample sizes could achieve low precision when ensuring factor stability, and may not have enough power to reject a false null hypothesis while being confident in this result (DeVellis, 2003).

Nevertheless, our results must be interpreted considering some limitations. First, the categorisation of instruments in terms of what aspect of mental well-being they assessed was based on our own assessment criteria with support from previous literature (Lara et al., 2019; Martín-María et al., 2017; Stone & Mackie, 2013), as most instruments did not explicitly state which aspects of well-being were assessed. In order

to deal with this limitation, instruments were classified by two independent reviewers and a third expert was consulted in case of doubt. Furthermore, some of the articles identified had missing information and none of the instruments tested cross-cultural validity, which means that its use with different samples would not be based on research. Additionally, subsequent validations and applications of the original instruments, and grey literature were excluded from this review, as the primary focus was on articles published in peer reviewed journals. The inclusion of grey literature does not guarantee the reduction of the publication bias (Martin, Pérez, Sacristán, & Álvarez, 2005) and could have restricted the validity of the results of this review, in terms of generalization of results. Related to the search strategy, date of production and limited descriptors could have affected the results obtained. Trying to cope with this issue, manual searches were performed to look for new instruments, being the case for SODA instrument (Fastame et al., 2019). Finally, the decision to include instruments that evaluated different aspects of mental well-being, (i.e. evaluative, hedonic, and eudaimonic well-being), allowed for an extensive general level of analysis, but limited the possibility to conduct thorough analyses, for example, a deep examination of the psychometric properties of items and subscales.

In conclusion, this review provides the first comprehensive synthesis of the existing instruments that assess mental well-being in the older population. The PWB (Reker & Wong, 1984), SCLSES (Salamon, 1988), HHS (Herth, 1991), LSITA (Barrett & Murk, 2006), MLS (Lee & Hong, 2017) and SODA (Fastame et al., 2019) were the most appropriated instruments. Instruments evaluating different aspects of mental well-being simultaneously and the existence of subscales in most of them are proof of the multi-dimensional nature of mental well-being. A mental well-being instrument specifically designed for the oldest old is needed.

References

- Barrett, A. J., & Murk, P. J. (2006). Life Satisfaction Index for the Third Age (LSITA): a measurement of successful aging.
- Carmel, S. (2017). The will-to-live scale: development, validation, and significance for elderly people. *Aging Ment Health, 21*(3), 289-296. doi: 10.1080/13607863.2015.1081149
- Carstensen, L. L., Fung, H. H., & Charles, S. T. (2003). Socioemotional selectivity theory and the regulation of emotion in the second half of life. *Motivation and emotion, 27*(2), 103-123.
- Closs, C., & Kempe, P. (1986). [A differential view and validation of the life-satisfaction construct: an analysis of proven procedures and recommendations for a methodologically based approach in measuring the dimensions of this construct]. *Z Gerontol, 19*(1), 47-55.
- Cohen-Mansfield, J., Shmotkin, D., Blumstein, Z., Shorek, A., Eyal, N., & Hazan, H. (2013). The old, old-old, and the oldest old: continuation or distinct categories? An examination of the relationship between age and changes in health, function, and wellbeing. *The International Journal of Aging and Human Development, 77*(1), 37-57.
- Cooke, P. J., Melchert, T. P., & Connor, K. (2016). Measuring well-being: A review of instruments. *The Counseling Psychologist, 44*(5), 730-757.
- Cresswell-Smith, J., Amaddeo, F., Donisi, V., Forsman, A. K., Kalseth, J., Martin-Maria, N., . . . Walhbeck, K. (2019). Determinants of multidimensional mental wellbeing in the oldest old: a rapid review. *Social psychiatry and psychiatric epidemiology, 54*(2), 135-144.
- Daaleman, T. P., Frey, B. B., Wallace, D., & Studenski, S. A. (2002). Spirituality Index of Well-Being Scale: development and testing of a new measure. *J Fam Pract, 51*(11), 952.
- Deaton, A. (2008). Income, health, and well-being around the world: Evidence from the Gallup World Poll. *Journal of Economic perspectives, 22*(2), 53-72.
- DeVellis, R. (2003). Factor analysis. Scale development, theory and applications. *Appl. Soc. Res. Method Ser, 26*, 10-137.
- Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49*(1), 71-75. doi: 10.1207/s15327752jpa4901_13
- Dronavalli, M., & Thompson, S. C. (2015). A systematic review of measurement tools of health and well-being for evaluating community-based interventions. *J Epidemiol Community Health, 69*(8), 805-815.
- Dunn, K. S. (2008). Development and psychometric testing of a new geriatric spiritual well-being scale. *Int J Older People Nurs, 3*(3), 161-169. doi: 10.1111/j.1748-3743.2007.00107.x
- Eakman, A. M., Carlson, M. E., & Clark, F. A. (2010). Factor structure, reliability, and convergent validity of the Engagement in Meaningful Activities Survey for older adults. *OTJR: Occupation, Participation and Health, 30*(3), 111-121. doi: 10.3928/15394492-20090518-01
- Fastame, M. C., Penna, M. P., & Hitchcott, P. K. (2019). SODA: a new questionnaire for the assessment of life satisfaction in late life span. *Aging Clinical and Experimental Research, 1*-19.

- Fox, K. R., Stathi, A., McKenna, J., & Davis, M. G. (2007). Physical activity and mental well-being in older people participating in the Better Ageing Project. *European journal of applied physiology*, 100(5), 591-602.
- Gorman, M. (1999). Development and the rights of older people.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., & Schünemann, H. J. (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *Bmj*, 336(7650), 924-926.
- He, W., Goodkind, D., & Kowal, P. R. (2016). *An aging world: 2015*: United States Census Bureau Washington, DC.
- Heo, J., Kim, J., Kim, B.-G., & Heo, S. (2014). Weekend experiences and subjective well-being of retired older adults. *American journal of health behavior*, 38(4), 598-604.
- Herth, K. (1991). Development and refinement of an instrument to measure hope. *Research and Theory for Nursing Practice*, 5(1), 39.
- Hyde, M., Wiggins, R. D., Higgs, P., & Blane, D. B. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging Ment Health*, 7(3), 186-194. doi: 10.1080/1360786031000101157
- Ingersoll-Dayton, B., Saengtienchai, C., Kespichayawattana, J., & Aunguroch, Y. (2004). Measuring psychological well-being: insights from Thai elders. *Gerontologist*, 44(5), 596-604.
- Jayawickreme, E., Forgeard, M. J., & Seligman, M. E. (2012). The engine of well-being. *Review of General Psychology*, 16(4), 327-342.
- Kobau, R., Snizek, J., Zack, M. M., Lucas, R. E., & Burns, A. (2010). Well-being assessment: An evaluation of well-being scales for public health and population estimates of well-being among US adults. *Applied Psychology: Health and Well-Being*, 2(3), 272-297.
- Kowal, P., & Dowd, J. E. (2001). Definition of an older person. Proposed working definition of an older person in Africa for the MDS Project. *World Health Organization, Geneva, doi*, 10(2.1), 5188.9286.
- Kozma, A., & Stones, M. J. (1980). The measurement of happiness: development of the Memorial University of Newfoundland Scale of Happiness (MUNSH). *J Gerontol*, 35(6), 906-912.
- Ku, P.-W., Fox, K. R., & McKenna, J. (2008). Assessing subjective well-being in Chinese older adults: The Chinese Aging Well Profile. *Social Indicators Research*, 87(3), 445-460. doi: 10.1007/s11205-007-9150-2
- Lara, E., Martín-María, N., Forsman, A. K., Cresswell-Smith, J., Donisi, V., Ådnanes, M., . . . Nyholm, L. (2019). Understanding the multi-dimensional mental well-being in late life: evidence from the perspective of the oldest old population. *Journal of Happiness Studies*, 1-20.
- Lawton, M. P. (1975). The Philadelphia geriatric center morale scale: A revision. *Journal of Gerontology*, 30(1), 85-89.
- Lawton, M. P., Kleban, M. H., Rajagopal, D., & Dean, J. (1992). Dimensions of affective experience in three age groups. *Psychology and Aging*, 7(2), 171.
- Lawton, M. P., Moss, M., Hoffman, C., Kleban, M. H., Ruckdeschel, K., & Winter, L. (2001). Valuation of life: A concept and a scale. *Journal of Aging and Health*, 13(1), 3-31.
- Lee, S. E., & Hong, G. R. (2017). [Development of the Meaning in Life Scale for Older Adults]. *J Korean Acad Nurs*, 47(1), 86-96. doi: 10.4040/jkan.2017.47.1.86
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., . . . Moher, D. (2009). The PRISMA statement for reporting systematic reviews and

- meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Annals of internal medicine*, 151(4), W-65-W-94.
- Liljas, A. E., Walters, K., Jovicic, A., Iliffe, S., Manthorpe, J., Goodman, C., & Kharicha, K. (2017). Strategies to improve engagement of 'hard to reach' older people in research on health promotion: a systematic review. *BMC Public Health*, 17(1), 349.
- Lindert, J., Bain, P. A., Kubzansky, L. D., & Stein, C. (2015). Well-being measurement and the WHO health policy Health 2010: systematic review of measurement scales. *The European Journal of Public Health*, 25(4), 731-740.
- Lou, W. Q., Chi, I., & Mjelde-Mossey, L. A. (2008). Development and validation of a life satisfaction scale for Chinese elders. *Int J Aging Hum Dev*, 67(2), 149-170. doi: 10.2190/AG.67.2.c
- Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research*, 46(2), 137-155. doi: 10.1023/A:1006824100041
- Makai, P., Brouwer, W. B., Koopmanschap, M. A., Stolk, E. A., & Nieboer, A. P. (2014). Quality of life instruments for economic evaluations in health and social care for older people: a systematic review. *Social science & medicine*, 102, 83-93.
- Marcoen, A., Van Cotthem, K., Billiet, K., & Beyers, W. (2002). Dimensies van subjectief welbevinden bij ouderen. = Dimensions of subjective well-being in elderly persons. *Tijdschrift voor Gerontologie en Geriatrie*, 33(4), 156-165.
- Martín-María, N., Miret, M., Caballero, F. F., Rico-Urbe, L. A., Steptoe, A., Chatterji, S., & Ayuso-Mateos, J. L. (2017). The impact of subjective well-being on mortality: a meta-analysis of longitudinal studies in the general population. *Psychosomatic medicine*, 79(5), 565-575.
- Martin, J. L. R., Pérez, V., Sacristán, M., & Álvarez, E. (2005). Is grey literature essential for a better control of publication bias in psychiatry? An example from three meta-analyses of schizophrenia. *European Psychiatry*, 20(8), 550-553.
- Mather, M., & Carstensen, L. L. (2005). Aging and motivated cognition: The positivity effect in attention and memory. *Trends in cognitive sciences*, 9(10), 496-502.
- McDowell, I. (2010). Measures of self-perceived well-being. *J Psychosom Res*, 69(1), 69-79.
- McMahan, E. A., & Estes, D. (2012). Age-related differences in lay conceptions of well-being and experienced well-being. *Journal of Happiness Studies*, 13(1), 79-101.
- McNeil, J. K., Stones, M. J., & Kozma, A. (1986). Subjective well-being in later life: Issues concerning measurement and prediction. *Social Indicators Research*, 18(1), 35-70.
- Meadow, H. L., Mentzer, J. T., Rahtz, D. R., & Sirgy, M. J. (1992). A life satisfaction measure based on judgment theory. *Social Indicators Research*, 26(1), 23-59. doi: 10.1007/BF00303824
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine*, 151(4), 264-269.
- Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of life research*, 27(5), 1171-1179.
- Mokkink, L. B., Prinsen, C. A., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2016). The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian journal of physical therapy*, 20(2), 105-113.

- Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., . . . De Vet, H. C. (2010). 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.
- Neugarten, B. L., Havighurst, R. J., & Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology*.
- Peat, J. (2001). *Health science research: A handbook of quantitative methods*: Sage.
- Poon, L. W., & Cohen-Mansfield, J. (2011). *Understanding well-being in the oldest old*: Cambridge University Press.
- Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research, 27*(5), 1147-1157.
- Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., . . . Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a “Core Outcome Set”—a practical guideline. *Trials, 17*(1), 449.
- Reijula, J., Rosendahl, T., Reijula, K., Roilas, P., Roilas, H., & Sepponen, R. (2009). A new method to assess perceived well-being among elderly people—a feasibility study. *BMC Geriatr, 9*(1), 55.
- Reker, G. T., & Wong, P. T. (1984). Psychological and physical well-being in the elderly: The Perceived Well-Being Scale (PWB). *Canadian Journal on Aging/La Revue canadienne du vieillissement, 3*(1), 23-32.
- Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring mental wellbeing among adolescents: a systematic review of instruments. *Journal of Child and Family Studies, 26*(9), 2349-2362.
- Ruffing-Rahal, M. A. (1991). Initial psychometric evaluation of a qualitative well-being measure: the Integration Inventory. *Health Values: The Journal of Health Behavior, Education & Promotion, 15*(2), 10-20.
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology, 52*(1), 141-166.
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology, 57*(6), 1069.
- Salamon, M. J. (1988). Clinical use of the Life Satisfaction in the Elderly Scale. *Clinical Gerontologist, 8*(1), 45-54.
- Salokangas, R. K., Joukamaa, M., & Mattila, V. (1988). Measurement of life satisfaction. Developing a life satisfaction scale. *Compr Gerontol B, 2*(2), 69-74.
- Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *Journal of personality and social psychology, 67*(6), 1063.
- Slivinske, L. R., Fitch, V. L., & Morawski, D. P. (1996). The Wellness Index: developing an instrument to assess elders' well-being. *Journal of Gerontological Social Work, 25*(3/4), 185-204.
- Step toe, A., Deaton, A., & Stone, A. A. (2015). Subjective wellbeing, health, and ageing. *The Lancet, 385*(9968), 640-648.
- Step toe, A., Demakakos, P., & de Oliveira, C. (2012). The psychological well-being, health and functioning of older people in England. *The dynamics of ageing, 98*.

- Stone, A. A., & Mackie, C. E. (2013). *Subjective well-being: Measuring happiness, suffering, and other dimensions of experience*: National Academies Press.
- Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M., Patrick, D. L., Alonso, J., . . . Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Quality of life research, 27*(5), 1159-1170.
- Thomson Reuters. (2013). EndNote X7: Thomson Reuters Philadelphia, PA.
- Torres, S. (1999). A culturally-relevant theoretical framework for the study of successful ageing. *Ageing & Society, 19*(1), 33-51.
- Tsang, K. L. V., Wong, P., & Lo, S. K. (2012). Assessing psychosocial well-being of adolescents: A systematic review of measuring instruments. *Child: care, health and development, 38*(5), 629-646.
- United Nations. (2019). World population prospects 2019: Highlights (ST/ESA/SER.A/423). Department of Economic and Social Affairs Population Division
- Unsanit, P., Sunsern, R., Kunsongkeit, W., O'Brien, M. E., & McMullen, P. C. (2012). Development and Evaluation of the Thai Spiritual Well-Being Assessment Tool for Elders with a Chronic Illness. *Pacific Rim International Journal of Nursing Research, 16*(1), 13-28.
- Zaslavsky, O., Rillamas-Sun, E., Woods, N. F., Cochrane, B. B., Stefanick, M. L., Tindle, H., . . . LaCroix, A. Z. (2014). Association of the selected dimensions of eudaimonic well-being with healthy survival to 85 years of age in older women. *International Psychogeriatrics, 26*(12), 2081-2091.

Appendix

Search strategies in PubMed, PsycINFO, Proquest Research Library, AgeLine, and CINHALL:

PubMed: ((validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)) in Title/Abstract and Humans [MeSH Terms].

PsycINFO: ((TI (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND TI (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND TI (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)) OR ((AB (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND AB (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND AB (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)). Without dissertations, books; only academic publications.

ProQuest Research Library: ((ab(enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”) AND ab(geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND ab(validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”)) OR ti((enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”) AND ti(geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND ti(validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”)). Academic articles in scientific journals.

AgeLine: ((TI (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND TI (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND TI (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)) OR ((AB (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND AB (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND AB (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)). Without dissertations, books; only academic publications.

CINAHL: ((TI (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND TI (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND TI (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)) OR ((AB (validation OR “factor analysis” OR validity OR reliability OR “psychometric properties” OR “factor structure” OR “cluster analysis” OR “principal component analysis”) AND AB (geriatric OR aging OR oldest OR older OR “old population” OR “old age” OR “very old” OR ageing OR elder OR elderly) AND AB (enjoy OR happ* OR optimis* OR “positive affect” OR “positive emotion” OR satisf* OR “well-being” OR wellbeing OR “self-realization” OR “personal growth” OR “purpose in life”)). Without dissertations, books; only academic publications.

Table 1. *Original mental well-being instruments and their principal characteristics*

<u>Instrument (acronym)</u> (author/s, year)	<u>Aspect of MWB</u>	<u>Age of participants (mean)</u>	<u>Sample size, gender (nation) Target population</u>	<u>Subscales/factors (number of items) (Response options)</u>	<u>Original language. Available translations</u>	<u>Interpretability</u>	<u>Feasibility (type and ease of application; cost of the instrument; length; completion time)</u>
*Revised Philadelphia Geriatric Center Morale Scale (RPGCMS) (Lawton, 1975)	Md	Not reported (mean = 72.6)	828 mw (United States) General population	-Agitation (6) -Attitude toward aging (5) -Lonely dissatisfaction (6) <i>Yes/No answer</i>	English. Chinese, Czech, Dutch, French, Japanese, Spanish, Swedish, Turkish	Total score ranges from 0 to 17 13+ represent high scores 10-12 mid-range scores <9 are low scores	Self- or interviewer-administered; available in the public domain; 17 items; approximately 10 minutes
Scale of Happiness of the Memorial University of Newfoundland (MUNSH) (Kozma & Stones, 1980)	Md	65-95 y (not reported)	598 mw (United States) Urban, rural and institutional residents	-Positive affect (5) -Negative affect (5) -General positive experience (7) -General negative experience (7) <i>Yes/No answer</i>	English. French, Spanish	Points for negative affect and experience items are subtracted for positive scores. The higher the total MUNSH score (maximum of 24 points), the higher the individual's happiness rating	Self-administered; available in the public domain; 24 items; approximately 20 minutes
*Perceived Well-Being Scale (PWB) (Reker & Wong, 1984)	Md	61-93 y (not reported)	238 mw (United States) General population	-Psychological well-being (6) -Physical well-being (8) <i>7-point Likert scale ranging from 1 (Strongly Agree) to 7 (Strongly Disagree)</i>	English. Arabic, Chinese, French, Turkish	A score of 7 on each item reflects a high level of well-being. Total score ranges from 6 to 42 for psychological well-being; 8 to 56 for physical well-being; 14 to 98 for general well-being	Self-administered; available in the public domain; 14 items; approximately 10 minutes
Satisfaction With Life Scale (SWLS)	Ev	Not reported (mean = 75.0)	53 mw (United States)	5 items	English. Arabic, Bosnian, Chinese, Czech, Danish,	Total score ranges from 5 to 35. A higher score indicates a higher life	Self-administered; available in the public

(Diener et al., 1985)			General population	7-point Likert scale ranging from 1 (Strongly Agree) to 7 (Strongly Disagree)	Dutch, Flemish, French, Georgian, German, Greek, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Khmer, Korean, Malay, Norwegian, Persian, Portuguese, Polish, Romanian, Russian, Serbian, Setswana, Spanish, Thai, Turkish, Urdu, Vietnamese	satisfaction. A value of 20 represents a neutral point on the scale	domain; 5 items; short administration time
Life-satisfaction construct (Closs & Kempe, 1986)	Ev	54-95 y (mean = 80.0)	457 w (Germany) Residents and retirement home population	-Social integration/ loneliness (11) -Satisfaction with life in old age (9) -Subjective somatic symptoms (11) -Tranquillity/ insecurity; concern (5) -Retrospective; congruence (3) <i>Yes/No answer</i>	German. Not available	Insufficient information. A higher score indicates a higher life satisfaction	Interviewer- administered; available in the public domain; 39 items; approximately 35 minutes
Life Satisfaction Scale (LSS-A) (Salokangas, Joukamaa, & Mattila, 1988)	Md	63+ y (not reported)	325 mw (Finland) General population	26 items (no information about their distribution) -Psychic balance -Assessment of past life -Present happiness <i>Yes/No answer</i>	English. Not available	Insufficient information. A higher score indicates a higher life satisfaction	Self-administered; available in the public domain; 26 items; approximately 20 minutes
Salamon-Conte Life Satisfaction	Md	55-90 y (not reported)	650 mw (United States) General population	-Pleasure in daily activities (25) -Meaningfulness of life (25) -Fit between desired & achieved goals (25) -Mood tone (25) -Self-concept (25)	English. Chinese, French, Italian, Persian	Insufficient information. A higher score indicate higher life satisfaction	Self-administered; not available in the public domain; 200 items; long administration time

in the Elderly Scale (SCLSES) (Salamon, 1988)				-Perceived health (25) -Financial security (25) -Social contacts (25) <i>5-point Likert-scale (ranging not reported)</i>			
*Six Scales of Psychological Well-Being (PWBS) (Ryff, 1989)	Eu	Not reported (mean = 75.0)	80 mw (United States) General population	-Self-acceptance (20) -Positive relation with others (20) -Autonomy (20) -Environmental mastery (20) -Purpose in life (20) -Personal growth (20) <i>6-point Likert scale ranging from 1 (Strongly Agree) to 6 (Strongly Disagree)</i>	English. Arabic, Chinese, Dutch, Filipino, French, Italian, Japanese, Persian, Romanian, Russian, Serbian, Setswana, Spanish, Swedish, Turkish, Urdu	Insufficient information. A higher score indicates a higher presence of each particular trait	Self-administered; 120 items; available in the public domain; administration time ranging from 45 to 75 minutes
*Herth Hope Scale (HHS) (Herth, 1991)	Eu	Study one: 62-92 y (mean = 72.2) Study two: 63-94 y (mean = 79.0)	Study one: 40 mw (United States) General population Study two: 75 mw (United States) Widowed	-Temporality and future (10) -Positive readiness and expectancy (10) -Interconnectedness (10) <i>Applies to me/ Does not apply to me answers</i>	English. Chinese, Dutch, German, Italian, Japanese, Korean, Norwegian, Persian, Portuguese, Spanish, Swedish, Thai	Total score ranges from 0 to 90. Higher scores represent greater hope	Self-administered; available in the public domain; 30 items; approximately 25 minutes
Integration Inventory (II) (Ruffing-Rahal, 1991)	Md	60-98 y (mean = 77.0)	156 mw (United States) General population	37 items (no information about their distribution) -Activity -Affirmation -Synthesis <i>6-point Likert scale ranging from 1 (Strongly Agree) to 6 (Strongly Disagree)</i>	English. Not available	Total instrument score is calculated by summing individual item scores (1-6). Scoring is reversed on a number of negatively worded items	Interviewer-administered; available in the public domain; 37 items; approximately 30 minutes
Congruity Life Satisfaction Scale	Ev	Study one: 60+ y	Study one: 752 mw	10 items	English. Not available	An overall general life satisfaction score is computed by summing	Self-administered; available in the public

(Meadow et al., 1992)		(mean = 70.5) Study two: 55+ y (mean =72.9)	Study two: 529 mw (United States) General population	6-point Likert scale varying from 1 (Very Dissatisfied) to 6 (Very Satisfied)		individual satisfaction scores (items 1 through 10) and dividing by the number of items (10)	domain; 10 items; approximately 5 minutes
Wellness Index (Slivinske, Fitch, & Morawski, 1996)	Md	62+ y (mean =73.4)	463 mw (United States) General population	-Physical health (12) -Morale (20) -Economic resources (10) -ADL-IADL (13) -Religiosity (11) -Social resources (13) 5-point Likert-scale (ranging not reported)	English. Not available	Insufficient information. A higher score indicates a higher wellness	Self- administered; available in the public domain; 79 items; long administration time
Subjective Happiness Scale (SHS) (Lyubomirsky & Lepper, 1999)	Eu	Not reported (mean = 69.5)	622 mw (United States) General population	4 items 7-point Likert scale ranging from 1 (Not at all) to 7 (A great deal)	English. Arabic, Chinese, French, German, Greek, Hungarian, Italian, Japanese, Malay, Portuguese, Romanian, Russian, Serbian, Slovak, Spanish, Turkish, Urdu	To score the scale, reverse code the 4 th item (i.e., turn a 7 into a 1, a 6 into 2, and so on), and compute the mean of the 4 items. Total score ranges from 1 to 7. Higher scores reflect greater happiness	Self- administered; available in the public domain; 4 items; short administration time
*Valuation Of Life Scale (VOL) (Lawton et al., 2001)	Eu	70+ y (not reported)	616 mw (United States) General population	-Hope (2) -Futurity (3) -Purpose (8) -Self-efficacy (2) -Perseverance (4) 5-point Likert scale ranging from 5 (Agree very strongly) to 1 (Disagree very strongly)	English. German, Japanese, Portuguese	Higher scores indicate higher valuation of life	Self- administered; available in the public domain; 19 items; approximately 15 minutes
Belgian Subjective Well-Being Scale	Md	60+ y (not reported)	366 mw (Belgium) General population	-Psychological (14) -Physical (10) -Social (9) -Material (10)	Dutch. Not available	No information available	Self- administered; available in the public domain; 56 items; long administration time

(Marcoen, Van Cotthem, Billiet, & Beyers, 2002)				-Cultural (6) -Existential (7) <i>7-point Likert scale ranging from 1 (Never) to 7 (Strongly agree)</i>			
Spirituality Index of Well-Being Scale (SIWB) (Daaleman et al., 2002)	Eu	Not reported (“Elderly individuals”)	277 mw (United States) Clinical population	-Self-efficacy (6) -Life scheme (6) <i>5-point Likert scale ranging from 1 (Strongly agree) to 5 (Strongly disagree)</i>	English. Chinese, Korean, Spanish	Total score ranges from 12 to 60. Higher scores indicate higher spiritual well-being	Self-administered; available in the public domain; 12 items; approximately 5 minutes
Control, Autonomy, Self-realization, Pleasure (CASP-19) (Hyde et al., 2003)	Eu	65-75 y (not reported)	286 mw (United Kingdom) General population	-Control (6) -Autonomy (5) -Self-realization (4) -Pleasure (4) <i>4-point Likert scale ranging from 0 (Never) to 3 (Often)</i>	English. Amharic, Arabic, Chinese, Czech, Lithuanian, Persian, Polish, Portuguese, Malay, Russian, Spanish, Turkish	Total score ranges from 0 to 57; 0 represents a complete absence of mental well-being, whereas 57 represents total satisfaction on all four domains	Self-administered; available in the public domain; 19 items; approximately 15 minutes
Thai Elders Psychological Well-Being measure (Ingersoll-Dayton et al., 2004)	Eu	60+ y (mean = 70.9)	460 mw (Thailand) General population	-Harmony (3) -Interdependence (3) -Respect (3) -Acceptance (3) -Enjoyment (3) <i>4-point Likert scale ranging from 1 (Strongly disagree) to 4 (Strongly agree)</i>	Thai. Not available	Total score ranges from 15 to 60. Higher scores indicate higher psychological well-being	Self-administered; available in the public domain; 15 items; approximately 10 minutes
*Life Satisfaction Index Third Age (LSITA) (Barrett & Murk, 2006)	Md	50+ y (not reported)	654 mw (United States) General population	35 items (no information about their distribution) -Zest vs. Apathy -Resolution and fortitude -Congruence of goals -Self-concept -Mood tone <i>6-point Likert scale ranging from 1 (Strongly disagree) to 6 (Strongly agree)</i>	English. Portuguese	Insufficient information. A higher score indicates a higher life satisfaction	Self-administered; available upon request; 35 items; approximately 30 minutes

Chinese Aging Well Profile (CAWP) (Ku et al., 2008)	Md	50+ y (mean = 62.1)	1419 mw (China) General population	-Physical (5) -Psychological (5) -Independence (5) -Learning & growth (4) -Material (4) -Environmental (4) -Social (4) <i>5-point Likert scale ranging from 1 (Strongly disagree) to 5 (Strongly agree)</i>	Chinese. Not available	Insufficient information. A higher score indicates a higher life satisfaction	Self- administered; not available in the public domain; 31 items; approximately 25 minutes
Geriatric Spiritual Well-Being Scale (GSWS) (Dunn, 2008)	Eu	61-100 y (mean = 74.2)	138 mw (United States) General population	-Affirmative self-appraisal (4) -Connectedness (4) -Altruistic benevolence (4) -Faith ways (4) <i>6-point Likert scale ranging from 1 (Strongly disagree) to 6 (Strongly agree)</i>	English. Lithuanian, Thai	Total score ranges from 16 to 96. Higher scores indicate higher spiritual well-being	Self- administered; available in the public domain; 16 items; approximately 10 minutes
Life Satisfaction Scale for Chinese Elders (LSS-C) (Lou et al., 2008)	Ev	60-94 y (mean = 69.7)	1502 mw (China) General population	8 items (no information about their distribution) -Instrumental needs of daily life (food/meals, finance, housing, job, health, transportation) -Social and relational needs (family relationships, responsibility, recreational activity, respect, intergenerational communication, friendship, partner, spirituality) <i>Satisfied/ Dissatisfied</i>	Chinese. Not available	A score of 7 or above (at or above the 70 th percentile) indicates a higher level of life satisfaction, and a score of 4 or below (at or below the 33 rd percentile) indicates a lower level of life satisfaction	Self- administered; not available in the public domain; 8 items; short administration time
*Con-Dis device for measuring	He	63-89 y (mean = 78.0)	10 mw (Finland) Resident population	Buttons of the Con-Dis device: -Happy -Neutral	Not reported	Values range from 1 (unhappy) to 3 (happy). Con-Dis is a countable device to measure an	Self- administered; developer company (Atmel corporation) is currently not active; variable period

perceived well-being (Reijula et al., 2009)				-Unhappy 420 individual markings in a 2-week period		elderly person's well-being	of follow-up (2w in this study)
Meaningful Activity Participation Assessment (MAPA) (Eakman et al., 2010)	Eu	65-100 y (mean = 80.5)	154 mw (United States) General population	28 items/ activities <i>For each activity: 7-point Likert scale ranging from 0 (Not at all) to 6 (Every day) and a 5-point scale from 0 (Not at all meaningful) to 4 (Extremely meaningful)</i>	English. Chinese, French	The total MAPA score involved the sum of the frequency rating multiplied by the meaning rating for each of the 28 items. Score ranges from 0 to 672, with higher scores indicating greater perceived meaningful activity participation	Self-administered; available in the public domain; 28 items; approximately 45 minutes
Thai Spiritual Well-Being for elders (TSWBATECI) (Unsanit et al., 2012)	Eu	60+ y (not reported)	2160 mw (Thailand) Clinical sample	-Acceptance of chronic illness (6) -Happiness in life (7) -Life equilibrium (5) -Passion for life (6) -Self-transcendence (5) -Optimistic personality (5) -A purpose in life (4) -Willingness to forgive (3) <i>4-point Likert scale ranging from 0 (Strongly disagree) to 4 (Strongly agree)</i>	Thai. Not available	Total score ranges from 0 to 164. Higher scores suggest a higher sense of spiritual well-being	Self-administered; available in the public domain; 41 items; approximately 35 minutes
Will To Live (WTL) (Carmel, 2017)	Eu	78-99 y (mean = 83.9)	868 mw (Israel) General population	5 items <i>6-point Likert scale ranging from 0 (Much stronger) to 5 (I have no will to live)</i>	English. Not available	The overall score is based on the average score of the responses to all 5 questions. Total score ranges from 0 to 5. A higher score indicates a higher will to life	Self-administered; available in the public domain; 5 items; short administration time

Meaning in Life Scale (MLS) (Lee & Hong, 2017)	Eu	65-90 y (mean = 75.0)	371 mw (Korea) General population	-Value of life (6) -Source of life (3) -Will to live (3) <i>4-point Likert scale ranging from 1 (Nothing) to 4 (Very much)</i>	Korean. Not available	Total score ranges from 12 to 48. A higher score indicates a higher meaning in life	Self-administered; available in the public domain; 12 items; approximately 5 to 10 minutes
*SODdisfazione dell'Anziano (SODA) (Fastame et al., 2019)	Md	60-98 y 60-69 y (mean = 63.4) 70-79 y (mean = 73.9) 80-89 y (mean = 83.9) 90+ y (mean = 91.9)	474 mw (Italy) General population 138 mw 197 mw 116 mw 23 mw	-SODA-health (8) -SODA-religious (2) -SODA-time (4) <i>10-point Likert scale ranging from 0 means (Not being satisfied at all) to 10 (Completely satisfied)</i>	Italian. Not available	Total score ranges from 0 to 140. A higher score indicates a higher life satisfaction	Self-administered; available in the public domain; 14 items; approximately 10 minutes

Note. Instruments are listed in chronological order. *Identified in the manual search.

Abbreviations: MWB (Mental well-being); Eu (Eudaimonic well-being); Ev (Evaluative well-being); He (Hedonic well-being); Md (Multi-dimensional well-being). Sample: y= year(s); w= women; mw= men and women

Table 2. *Quality appraisal of measurement properties according to the original instrument*

<u>Instrument</u>	<u>Content validity</u>	<u>Structural validity</u>	<u>Hypothesis testing for construct validity</u>	<u>Criterion validity</u>	<u>Internal consistency</u>	<u>Test-retest reliability</u>	<u>Quality level of evidence</u>
Evaluative well-being							
SWLS	+	?	?	-	-	+	Low
Life-satisfaction construct	+	+	?	?	-	?	Very low
Congruity Life Satisfaction Scale	+	?	+	-	+	?	Low
LSS-C	+	+	?	-	+	+	Moderate
Eudaimonic well-being							
PWBS	+	+	?	-	+	+	Moderate
HHS	+	+	?	+	+	+	High
SHS	+	?	?	+	+	-	Low
VOL	+	+	+	-	+	?	Moderate
SIWB	-	?	?	-	+	?	Very Low
CASP-19	+	+	?	-	-	?	Low
Thai Elders Psychological Well-Being measure	+	+	?	-	+	-	Low
GSWS	+	-	+	-	+	-	Low
MAPA	+	?	+	-	+	+	Moderate
TSWBATECI	+	+	?	?	+	?	Low
WTL	+	+	+	-	+	?	Moderate
MLS	+	+	+	+	+	+	High
Hedonic well-being							
Con-Dis device for perceived well-being	+	?	?	+	?	?	Low
Multi-dimensional							
RPCGMS	-	-	?	+	+	+	Low
MUNSH	+	?	?	+	+	+	Moderate
PWB	+	-	+	+	+	+	High
LSS-A	+	?	?	-	+	?	Low
SCLSES	+	+	+	+	-	+	High
II	+	?	?	-	+	?	Low
Wellness Index	+	-	+	-	+	-	Low
Belgian Subjective Well-Being scale	+	?	?	?	+	?	Low
LSITA	+	+	+	+	+	?	High
CAWP	+	+	?	-	+	-	Low
SODA	+	+	+	-	+	+	High

Note. The results were classified as positive (+), indeterminate (?), or negative (-) according to the quality criteria for each measurement property. Dark green indicates the instruments with a high level of quality evidence, whereas light green points out the ones that reached a moderate level.

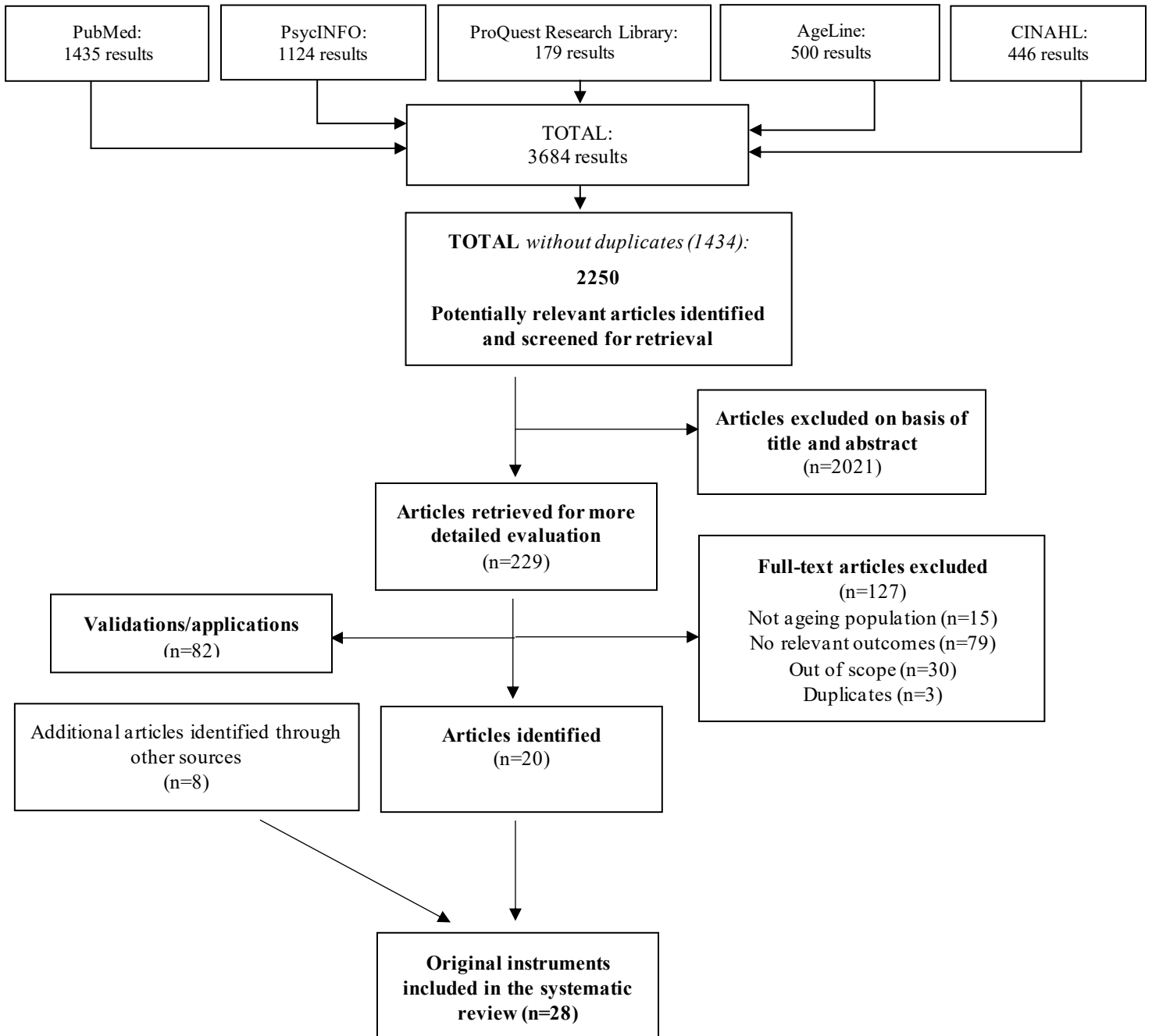


Figure 1. Results obtained from the five search strategies and flow diagram of the systematic review