DIMENSIONALITY ASSESSMENT OF THE MEASURE OF MUNDANE MEANING

Olga Andrew

July 2023

Research submitted in partial fulfilment of the requirements for the degree of Doctor in Clinical Psychology (DClinPsy), Royal Holloway, University of London.

Acknowledgements

I would like to express my gratitude to the people who have played an important role in the completion of this thesis. First, I want to thank my academic supervisor, Dr Gary Brown, for his knowledge and support throughout this project. I am also grateful for Ben, whose encouragement and love (and everything else!) have got me through this difficult time. Your support has let me to focus on this project, even as we have gone through the challenges life has thrown at us! I also want to thank Hero. Having a best friend to help with statistics and R has been invaluable. Our late-night support sessions got me through the final push – thank you! I also want to acknowledge Spot, whose naps on my lap have provided much comfort throughout this project. Your little wagging tail has made it so much easier to settle down and encouraged me to take very regular breaks! Lastly, I want to thank E. Your presence has given me such happiness and hope to keep going. I cannot wait to see what the future holds!

To Dr Gary Brown, Ben, Hero, Spot, and E, I am grateful for your support and inspiration. This thesis would not have been possible without each of you.

With heartfelt thanks,

Olga

Table of Contents

Lay summary

Empirical Study

The Measure of Mundane Meaning is a survey that measures people's experience of meaning in life. People with trauma came up with the questions. Trauma can make you think about meaning. This can give a new perspective. This makes this survey different from other surveys.

The survey had 36 questions. For each questions people rate how much they agree with it. There are seven options to choose from. These range from "not at all true of me" to "completely true of me." Here is an example question: "I know what's important in my daily life."

We do not know if this survey is good. A good survey measures what it says it will measure. It can be hard to know which questions are good. A way to find out more is to look at people's answers.

Questions in surveys can be similar. These similar questions can form a sub-group. There might be different sub-groups of meaning in life. We had thought about four topics. One is having a plan or mission. Another one is having a life story that makes sense. One is making sense of things that happen to you. The last one is knowing what to do. We thought these were the topics of the survey.

We wanted to check if the survey had these four topics. We gave the survey to 893 who were made up of a general sample. These were different groups of people:

- 192 mothers and 68 women following a miscarriage
- 74 adults with long-term pain
- 120 unemployed or 42 recently re-employed adults
- 267 adults who were recruited through an online platform
- 160 students or recent graduates

Then we used a computer program. The first thing the computer program checked was whether all questions were of good quality. If they were poor quality, they would join different groups each time. We removed these questions which seemed random. At this stage we had 24 questions that were good quality. Then the computer program checked if people answered some questions similarly. The computer program can say how many groups of questions there are. There are many ways of doing this. Some programs are better at knowing how many groups there are. We used a new technique which is called “Exploratory Graph Analysis” which has been found to be better than some others.

We found that some questions grouped together. There were four topics/subgroups. Then we looked at the questions for each topic. This helped us know the topic for each group. We found four topics we predicted. These were: having a plan or mission, making sense of things that happen to you and knowing what to do. A life story that makes sense was also one of the topics.

Our guess was that there would be four groups. There were four groups, which were the same as expected. We now have trust that we know what this survey measures. This can help people who are missing meaning in life. This survey tells us what parts of meaning in life a person does not have. For example, we can tell if people need help with making goals. They might also need help with making sense of what has happened. Using this survey can help people know what to focus on.

One thing about this study that is not good is that we used many different groups of people. For example, some people had trauma and some people did not. This means that we do not know how different people respond on the survey. Maybe some groups of people fill it in differently. We do not know this. In a way having many different people is also good. This is because we know that different people can use the survey.

In summary, we studied a survey of meaning in life. People with trauma came up with the questions on this survey. We found that the questions formed certain groups which people of a general population sample answered similarly. These four groups were: having a plan or mission, making sense of things that happen to you and knowing what to do, and making sense of your life. We can use this survey to help people who want to understand themselves.

Systematic Review

Many measures of meaning in life exist. In 2012 a collection of all previously developed surveys was published. Some surveys were of better quality, and some were worse quality. The aim of this study was to collect all meaning in life surveys which were developed since 2012.

We used Medline and Psycinfo, which collect published articles. All studies which were about the development of surveys about meaning in life were included. Only articles which were written in English and had included adult participants were chosen. Also, only articles which had been evaluated by other professionals (peer-reviewed) were included. To look at the quality of the surveys, the Cosmin framework was used. It is recommended to use the Cosmin framework when evaluating the quality of surveys.

We included ten studies. Each study scored badly using the Cosmin framework. This means that the studies were of poor quality. Often, this was because the surveys had not been tried with the people who would fill them out. This means it is unclear if they measure what they intent to measure. This means that we were not able to recommend a specific survey to be used when measuring meaning in life.

Dimensionality assessment of the Measure of Mundane Meaning

Abstract

Meaning in life is an important aspect of psychological wellbeing. The Measure of Mundane Meaning (MMM) measures the presence of meaning in life and is unique in its development among participants with experience of trauma. The MMM was hypothesised to comprise of four factors including sense of purpose, high-level action identification, integration of circumstance, and coherence of self-narrative and the aim of the current study is to conduct a dimensionality assessment of the MMM in a general population sample.

The study utilised a novel psychometric technique, Exploratory Graph Analysis, to analyse the 36-item MMM. These 893 participants were a combination of clinical and non-clinical samples.

Redundancy was assessed using Unique variable analysis (UVA) and the stability of the items was assessed. Random intercept EGA (riEGA), which is a modified version of EGA that can account for wording effects was utilised. In this process, twelve items were removed, and the remaining 24 items formed four dimensions. Using confirmatory factor analysis, this model was found to exhibit good fit, and a multidimensional model was favoured. The final sample consisted of four dimensions, which represented the four hypothesised dimensions. The relationship between the MMM and other measures of meaning, depression, and underlying assumptions was evaluated to assess convergent validity along with developing an understanding of the relationship between the MMM and other underlying constructs.

This validation of the MMM provides a measure of meaning in life which corresponds to an underlying theoretical structure of the construct. This can enable more precise measurement of meaning in life, in both research and for clinical purposes.

Dimensionality assessment of the Measure of Mundane Meaning

Experiencing meaning or purpose regarding one's existence is an essential part of being a person (Frankl, 2004).

While not explicitly part of the diagnostic framework of mental health conditions, meaning in life is an important aspect of psychological wellbeing. Reduced meaning in life is related to an increase in psychological distress (Li et al., 2019). Individuals who experience mental health problems generally report lower meaning in life (Kleftaras & Parras, 2012) and meaning in life is a protective factor for suicide risk (Marco et al., 2016). This suggests meaning in life functions as a transdiagnostic construct that is relevant to individuals who experience mental health conditions.

Considering the relationship between mental health and meaning in life have led some to suggest that, rather than focusing on clinical recovery, the aim of clinical interventions should be to encourage meaning (van Weeghel et al., 2019). The transdiagnostic construct of meaning is equated with 'personal recovery' within the context of recovery from mental health difficulties. In this context, 'personal recovery' is understood as the attainment of meaning along with the potential continued experience of clinical symptoms of mental health conditions (Slade, 2009). This perspective takes a broader view of the impact of clinical services to consider how meaning in life can be encouraged. The Department of Health, which shapes the direction of mental health services within the UK, have placed an increasing emphasis on personal recovery over the past decade (Department of Health, 2011). Despite this, the largest providers of psychological therapy within the NHS do not take meaning into account when evaluating their effectiveness (National Collaborating Centre for Mental Health, 2018). This is probably at least partly due to the lack of valid measures of meaning in life which would enable services to evaluate their interventions against their impact on meaning.

At the international level, the Organisation for Economic Co-operation and Development (OECD, 2013) encourages its member states to include subjective wellbeing as a national indicator. In addition to evaluative and experiential wellbeing, they recommend eudaimonia is included, which has been conceptualised as meaning in life (OECD, 2013). Martela and Ryan (2023) note that while there is consensus on how to conceptualise and measure evaluative (life satisfaction) and experiential (positive and negative affect) wellbeing, this is not the case for eudaimonia/meaning in life.

Evaluating meaning in life in more detail can enable standardised measurement of meaning in life amongst individuals, within clinical services, and on a national level within our society. This would enable us to make comparisons about the effect that different experiences have on meaning in life, to compare changes in meaning in life within individuals, and assess national patterns. However, there are many challenges in measuring meaning in life. One of these is the multiple ways meaning is defined, Baumeister (1991, p. 15) defined meaning in life as the "mental representation of possible relationships among things, events, and relationships. Thus, meaning connects things". This definition suggests our lives become meaningful when patterns and connections are experienced. It may be necessary to see a pattern in the past to experience a sense of direction in the present and future. Meaning in life has also been defined as "the sense that our lives matter, that they make sense, and that they are more than the sum of our seconds, days, and years" (Steger, 2012, p. 65). This suggests that meaning in life is more than simply our experiences, and that these experiences are more "than the sum of their parts". How we understand our lives adds to its meaning, beyond what is experienced. Meaning in life, this definition seems to suggest, is something that cannot be observed from the outside, rather it is 'felt.' This definition would suggest difficulties in measuring meaning in life, as it relies on a person's internal experience. A further definition of meaning in life suggests that "lives may be experienced as meaningful when they are felt to have significance

beyond the trivial or momentary, to have purpose, or to have a coherence that transcends chaos" (King et al., 2006, p. 180). The above definitions are not an exhaustive list of definitions of meaning in life, instead they are an attempt to illustrate the multiplicity of definitions. Each definition seeks to describe the experience of meaning in life. However, these definitions all point to different ways of understanding meaning in life and thus fail to identify how meaning in life can be delineated from other similar constructs.

Researchers have argued that since there is not a clear definition of meaning in life, this has led to a weak shared understanding of the concept (Leontiev, 2013). On one hand, it is possible that including multiple definitions means our understanding of meaning in life becomes too broad. As we have seen above, all definitions provide a slightly different understanding of the concept. Definitions need to be specific to be useful, otherwise we cannot delineate them from other similar constructs. Also, it is not clear whether definitions of meaning in life are using different words to express the same construct which people instinctively understand, or whether the term is being used to refer to different constructs. Given that the definitions introduced for meaning in life area so varied, the latter appears more likely. The lack of an agreed-upon definition of meaning in life has negatively impacted our ability to measure the experience of meaning within individuals.

A further challenge to measuring meaning in life is a lack of an agreed upon structure of the construct of meaning in life. One way of ensuring the specificity of meaning in life is by defining it through its structure; making sure that there is clarity about the underlying dimensions which make up meaning in life. Due to the complexity of meaning in life as a concept, its meaning may be made up of multiple different aspects rather than a simple entity. But the lack of a clear definition means there is no consensus about whether meaning in life refers to a unidimensional construct, or whether, as some researchers have suggested, it consists of multiple dimensions (Martela & Steger, 2016). Further, if we assume that the concept of

meaning in life has a multi-dimensional structure, which underlying constructs make up that structure have not been agreed upon. These constructs would need to be determined for meaning in life to be reliably measured.

Despite the conceptual challenges of defining meaning in life, a variety of theories of underlying dimensions of meaning of life have been suggested (Martela & Steger, 2016). There are many examples of writers who proposed models of meaning in life which are based on defining it through its underlying structure. The first is Steger and colleagues (2006) who proposed a model of meaning in life which encompasses two dimensions. These dimensions are a) the significance felt regarding one's existence and b) the sense made of oneself and one's life (Steger et al., 2006). Another theory suggests that meaning in life consists of coherence, purpose, and significance (King & Hicks, 2021). The inclusion of significance is consistent with Steger and colleagues' definition, and coherence has a similar meaning to 'the sense made of oneself and one's life. However, purpose is a distinct dimension in King & Hicks' definition. George and Park (2016) proposed a similar model which conceptualises meaning in life as comprehension, purpose and mattering. The inclusion of purpose is consistent with King & Hicks, and there are also clear similarities between both 'comprehension' and 'coherence', and 'significance' and 'mattering'. This suggests that King & Hicks' definition and George & Park's definition are conceptually equivalent, in that they suggest a three-factor structure consisting of purpose, significance/mattering, and coherence/comprehension. Attempts to define meaning of life by considering the underlying factors which make up the concept, provide a deeper understanding of meaning in life, while also giving a framework for measuring it. Once the underlying structure of meaning in life has been identified, we can utilise this understanding to develop tools that evaluate and quantify the experience of meaning in life, considering each factor in turn.

Many measures of meaning in life exist (Brandstätter et al., 2012). The Meaning in Life Questionnaire (MLQ) developed by Steger and colleagues (2006) is the most prevalent and validated measure of meaning in life. In the development of this questionnaire, the structure of meaning in life was determined as consisting of significance and coherence/comprehension, as it was based on Steger and colleague's definition from the same study. This way of defining meaning in life suggests meaning in life has a two-factor structure, and the MLQ does indeed have a two-factor structure and is therefore made up of two subscales (Brandstätter et al., 2012). However, this two-factor structure is not related to the dimensions of significance and coherence/comprehension (Steger et al., 2006). Instead, the questionnaire is made up of two subscales: search and presence of meaning in life. These two subscales (search and presence) do not relate to the two-factor model that includes significance and coherence/comprehension. Therefore, the factor structure of this questionnaire does not relate to our understanding of the structure of meaning in life, or indeed to that of Steger and colleagues' own study. Moreover, studies such as King & Hicks' which suggest a three-factor model of meaning in life, including significance, purpose, and coherence, is now commonly used (King & Hicks, 2021). Therefore, the dominant understanding of meaning in life in the current literature is at odds with the prevailing way of measuring meaning in life – the MLQ – which is based on a different understanding. This means that there is no measure of meaning in life which corresponds to an up-to-date understanding of the concept, suggesting that there is no measure which is robust enough to be reliable, given our current understanding.

A measure has now been developed with a definition of meaning in life which is based on understanding its underlying structure. The Measure of Mundane Meaning (MMM) was developed to assess levels of personal meaning amongst individuals who have experienced a loss of meaning, such as among victims of trauma (Brown et al., 2008). It is hard to specify the structure of meaning in life a priori. Therefore, one way to further our understanding of

meaning in life is to evaluate meaning in life among those who have had their meaning in life threatened and to identify themes in their experiences. In contrast to other measures and theoretical constructs as described above, the MMM has been mapped onto four dimensions related to meaning in life (Brown et al., 2008). This four-factor structure includes coherence of self-narrative; integration of circumstances; high–level action identification and sense of purpose. Coherence and purpose are both factors which are included in prevailing definitions such King & Hicks' (2021) and were therefore included in the proposed structure of the MMM. Purpose was considered as 'goal-directed activity' which is equivalent to the way it was used in previous definitions and is generally considered an important aspect of meaning (Hill et al., 2015). In the development of the MMM, coherence of self-narrative was understood as a combination of self-narratives and the construct of 'Sense of Coherence'- the idea that an ability to construct a continuous narrative regarding one's experiences is an important part of experiencing meaning (Antonovsky, 1979).

The two new factors that are unique to the MMM are integration of circumstances and high-level action identification. 'Integration of circumstances' is linked to the coherence of self-narrative in that it relates to the ability to make sense of experiences that challenge assumptions (such as trauma) regarding our actions and what happens to us (Brown et al., 2008). However, while coherence of self-narrative suggests an ability to make sense of both our lives as a whole, integration of circumstance relates to the ability to make sense of specific traumatic experiences. The notion of the importance of coherence in our understanding of trauma was suggested by Foa and Kozak (1989) who argued that among people who develop PTSD, the significance of the threat of the event determines their level of distress. They noted that it was the meaning of the event that changed a person's experience of it, rather than the actual level of threat, or even the experience itself. Based on this theory, integration of circumstance was considered to be an important part of meaning structure among those who

have undergone trauma. The other new factor was high–level action identification, which relates to the experience of the meaning of our actions and examines how our everyday actions, such as travelling to work or cooking dinner, relate to a higher level of meaning in our lives (Watkins, 2011). When people who experience trauma develop PTSD a change in abstraction level has been found, as people who develop PTSD are more likely to evaluate the traumatic event on a higher level of abstraction (Watkins, 2011). Based on this research, high-level action identification was considered an important aspect of understanding meaning and therefore the structure of the MMM. These four aspects which are suggested to determine meaning in life provide a comprehensive view of meaning in life. To date, no empirical evidence for this theoretical structure has been determined, which is a necessary next step in the validation of the MMM.

The theoretical structure of the MMM shares similarities with other suggested structures but also provides novel contributions. Both coherence and purpose are suggested by other researchers (e.g. Steger et al., 2006; King & Hicks, 2021; George & Park, 2016). However, high-level action identification and integration of circumstance are novel contributions of the MMM. This difference in definition may be due to the fact that the MMM was developed to consider participants who have experienced trauma, which reveal to a person what is meaning to them. Integration of circumstances and high-level action identification may be necessary to experience meaning in life when one has experienced a threat of loss of meaning. These experiences may not be captured in other measures developed with non-clinical populations and the MMM therefore provides a unique perspective on meaning in life.

Above, the hypothetical dimensionality of the MMM has been presented. However, whether the MMM maps onto these four theoretical dimensions needs to be assessed within a large sample. When the MMM was developed, initial validation took place in a small sample and a dimensionality assessment was not possible (Brown et al., 2008). Further, this will allow

us to consider which items of the questionnaire that relates to each dimension. This process will allow us to determine the construct validity of the MMM.

The use of dimensionality assessments is particularly apt in this context following the development of new methods based on psychometric network models. Network approaches, compared to latent variable models do not rely on an assumption of causality, they only focus on interrelations between variables. Schmittman and colleagues (2013) argue latent variable models neglect the relationship between the observed variables themselves. From the network perspective, a construct is theorised to consist of variables which influence each other, with structure emerging out of their inter-relationships. Indicators are considered "autonomous causal entities in a network of dynamical systems" (Schmittman et al., 2013, p. 47). Such mutualism models encourage a focus on the direct relationships between observed variables (rather than between attributes and items). The observable variables do not measure a construct, instead they are part of it. Christensen and colleagues (2020c) argue that the psychometric network perspective aligns with our understanding of psychological traits as complex systems better than latent variable models. Psychometric network analysis does not aim to simplify constructs to the extent that their dynamic nature is ignored and does not assume clarity of constructs which do not exist (Christensen et al., 2020c). Therefore, this approach is appropriate to use in the assessment of the MMM.

In addition to conducting a dimensionality assessment of the MMM, a further aim of this study is to examine the relationship between the MMM and measures which tap into related constructs. Firstly, the association between the MMM and the MLQ (which is, as noted above, the most used and validated measure of meaning in life) will be evaluated to determine convergent validity of the MMM. With a similar aim, the relationship between the MMM and the meaningfulness subscales of the World Assumptions Scale (Janoff-Bulman, 1992) will be determined. It has been suggested that beliefs about oneself and the world impact experiences

of meaning (Park, 2010). Therefore, evaluating this relationship will further expand knowledge about how meaning in life relates to assumptions people hold.

The MMM and its relationship with depression will also be evaluated, as operationalised by the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983). An inverse relationship between depression and meaning in life has been found (Steger et al., 2006). By examining this relationship, the aim is to further develop an understanding of the relationship between meaning in life and depression.

In summary, the aim of this study was to determine the dimensionality and hence continue the process of establishing the validity of the MMM, a measure of meaning in life developed from a perspective of loss of meaning in life, within a general population sample. A four-dimensional theoretical dimensionality has been suggested and the result will be compared against these proposed structures. A novel psychometric technique, EGA, will be used to determine the dimensionality and structure of the underlying construct.

Methods

Participants

A cross-sectional design was utilised for archival analysis of previously published and unpublished data. The responses were collected over a period starting from 2008. The total sample consisted of 893 participants. A sample size of over 500 was associated with high levels of accuracy when conducting an EGA in a simulation sample (Golino & Epskamp, 2017).

The sample consisted of adult participants. This data was a combination of different studies and collection points which were combined for analysis. The samples were made up a population-based sample of both clinical and non-clinical samples from varying backgrounds.

One sub-sample consisted of 192 first-time mothers following the birth of their first child and 68 women following a miscarriage. Participants were recruited through local groups for new mothers and those who have experienced miscarriages in Greater London and Cambridge,

UK. The women completed the MMM on the internet after providing informed consent and agreeing to take part in the study.

A further sub-sample consisted of 74 adults experiencing chronic pain. The participants were recruited by clinicians in Greater London, UK. After they were provided information about the study and gave informed consent, participants were sent a battery of questionnaires including the MMM along with pre-paid envelopes. Of the 220 people who agreed to take part 33.7% returned the questionnaires by post.

The MMM was also collected from a group of 162 individuals who were either unemployed (N=120) or recently re-employed (N=42). Half of participants were based in London and half in other parts of the country. Participants were recruited through online advertising, posters in unemployment services and targeted emails. After providing informed consent they either filled out the questionnaire online or on paper depending on preference.

As part of a larger study 267 adults were recruited through the online platform http://www.prolific.co. Participants completed the study, which included the MMM, online after providing informed consent. They were provided £3.35 in compensation for completing the study.

As part of an MSc dissertation, the MMM was collected from a group of 160 students who currently attended or had recently attended a university in the UK. Participants were recruited through social media sites and the questionnaire was completed online after they provided informed consent. A high level of English was required and any responses from those who had accessed treatment for their mental health were excluded.

Measures

Measure of Mundane Meaning

The Measure of Mundane Meaning is a self-report questionnaire which aims to measure meaning in life. It includes 36 questions such as “I don’t have to think very hard about what I

need to do from moment to moment.“ The questionnaire is included in Appendix A. Preliminary psychometric testing of the MMM indicated a Cronbach’s alpha of .96 (Brown et al., 2008). A negative association with scores on the MMM and measures of anxiety and depression on the Depression Anxiety Stress Scales was found (Lovibond & Lovibond, 1995). This negative association with measures of symptomatic distress indicates validity of the questionnaire. Also, associations with subscales of the World Assumptions Scale related to benevolence of the world and people, and beliefs in a just world were significant in a preliminary sample (Janoff–Bulman, 1992). This relationship between holding these views and the MMM further suggested its validity.

Meaning in Life Questionnaire

The Meaning in Life questionnaire (MLQ) is a measure of the presence and search of meaning in life (Steger et al., 2006). Each subscale consists of five self-report items rated on a scale from one to seven. Both subscales have been found to exhibit good internal validity, with coefficient alpha of the presence subscale ranging from .81-.86, and the search subscale 0.84-.92 (Steger et al., 2006). Also, the scales exhibited good stability, as evidenced by test-retest stability coefficients for the presence subscale of 0.73 and the search subscale of 0.73 (Steger et al., 2006).

Hospital Anxiety and Depression Scale

The Hospital Anxiety and Depression Scale (HADS) is a screening tool for symptoms of depression and generalised anxiety disorder (Zigmond & Snaith, 1983). The depression subscale of the HADS (HADS-D) consists of seven questions related to depression which are scored on a Likert-style scale ranging from zero to three, enabling a total score of 0-21. The HADS has shown to exhibit good sensitivity and specificity in medical populations, in which it was developed (Wu et al., 2021). In a review of a combination of medical and community samples the internal consistency of the HADS-D was found to range from .67 to .90, with the

mean value of Cronbach's alpha reported at .82 (Bjelland et al., 2002). Furthermore, the HADS-D demonstrated good discrimination in a healthy sample, as suggested by the AUC (Area under the Curve) of 0.84 (Kjaergaard et al., 2014).

World Assumptions Scale

The World Assumptions Scale (WAS) is a 32–item scale assessing a person's assumptions regarding benevolence of the world, meaningfulness, and worthiness of the self (Janoff–Bulman, 1992). An eight-factor structure has been validated consisting of the following subscales: benevolence of the world (BW), benevolence of people (BP), justice (J), control (C), randomness (R), self–worth (SW), self–control (SC), and luck (L) with Cronbach's alpha of the subscales ranging from .68 to .84 (van Buggen et al., 2018).

Analytic strategy

The analysis was conducted within the R package, using EGAnet (Golino et al., 2021).

To assess the underlying structure of the MMM, a network approach was used to estimate the structure by creating a visual representation of the network structure, which indicates how items on the MMM are related to each other and whether communities between certain items form. Network models are visualised using a Graphical Gaussian Model (Lauritzen, 1996) with items as nodes and the empirical correlation between items as edges. This graph is created by nodes (where each node represents a variable) and links between these nodes (which visually represent the strength of the relationship between the variables by closeness between the variables or thickness of the links). Each edge represents a partial correlation coefficient between two variables after conditioning on all variables within the network (Golino & Epskamp, 2017). Compared to Bayesian network models, undirected network models allow a two-way relationship where there is no direction identified between the nodes, in other words, both nodes can influence each other. Network approaches interpret variables as interacting with each other in a mutually reinforcing network (Borsboom &

Cramer, 2013). Rather than being determined by a common latent variable, these variables are seen to reciprocally reinforce each other. However, close clusters of variables have been suggested to be representative of underlying latent variables and are indeed statistically equivalent to latent factors (Golino & Epskamp, 2017). This visual interpretation of the network provides both the structure and the dynamics of the network. The last part of the interpretation of the measure is the psychological meaning of the underlying dimensions.

Multiple sources of information were combined and synthesised to determine and evaluate the likely psychological meaning of the clusters. This included analysing the content of each item, network loadings, along with the relationship between clusters. Also, the proposed structure was compared with the theoretical framework of the MMM. Such approaches can provide insight into the meaning of the clusters, but ultimately requires the researcher to make a subjective judgment regarding the psychological meaning of the clusters based on a synthesis of the available information.

Redundancy

Prior to computing the graphical representation of the items, an assessment of redundancy was necessary. An assumption in network psychometrics is that variables are causally autonomous (Cramer et al., 2012) and therefore, redundancy violates assumptions of network psychometrics. It also produces further problems in the analysis of data, such as when a factor is formed by redundant variables and therefore misrepresenting network loadings by estimating covariance only between the redundant items. Unique variable analysis is a method developed within network psychometrics to evaluate local dependence (Christensen et al., 2020b). If items are found redundant with each other, the researcher can either combine them into one variable or remove the redundant variables. UVA has been found to be effective and supports the stability of the internal structure of a questionnaire (Christensen et al., 2020b). After assessing all variables as non-redundant, EGA can be evaluated.

Dimensionality

EGA consists of two steps, firstly by estimating a network and then identifying clusters using a community detection algorithm (Golino & Epskamp, 2017). The structure of the network is estimated by a Graphical Gaussian Model (GGM) by calculating the inverse of the variance-covariance matrix of the sample. However, the GGM often develops unstable parameters and spurious correlations which can impact the edges and cause large standard errors. To avoid this, a penalized maximum likelihood estimate, such as the graphical least absolute shrinkage and selection operator (GLASSO) is used, which guards against overfitting (Golino & Epskamp, 2017). Using this approach, relationships that are close to zero are modelled as zero, which leads to a sparse network. A tuning parameter needs to be used to retrieve the true network structure. For EGA, the extended Bayesian information criterion (EBIC) is used. All variables which lack a direct relationship are conditionally independent, in that they have no relationship after conditioning on all other variables in the network.

After the network is estimated, a community detection algorithm is used to find clusters in the network. When EGA was initially developed, Walktrap was used as the default community detection algorithm to determine clusters in the network (Golino & Epskamp, 2017). However, in a simulation study published since then, combining GLASSO with the community detection algorithm Louvain was found to increase general accuracy and reduce bias compared to using Walktrap (Christensen et al., 2020a). Therefore, EGA will be combined with the Louvain community detection algorithm (Blondel et al., 2008). Louvain community detection algorithm divides the network into clusters which maximise the modularity (quality of the clusters). Modularity is determined by the numbers of edges that exist within a subgroup compared to outside of this subgroup. EGA combined with the

Louvain community detection algorithm compute the number of dimensions, and their composition, along with a visual representation of the relationship between each item.

A concern regarding the accuracy of EGA is the impact that wordings effects may have. Wording effects are defined as "respondents' differential response style to positively and negatively worded items" (Kam, 2018, p.574). Negatively and positively scored items are often included in questionnaires to measure a construct, and the MMM consists of both negatively and positively worded items. In factor analyses, differently worded items may load onto different factors, suggesting the semantics impact the factor analysis, rather than the underlying construct (Kam, 2018). Since wording effects have been found to reduce the accuracy of EGA, Garcia-Padina and colleagues (2022) have developed a modified EGA approach, called random intercept EGA (riEGA). riEGA was found to increase the accuracy of EGA in a sample with wording effects but did not reduce the accuracy in a sample with no wording effect present (Garcia-Padina et al., 2022). This suggests that riEGA is the most appropriate method to use in this study, as the MMM consists of both negatively and positively worded items.

Structural Consistency and Replicability

EGA has shown accuracy in estimating dimensions compared to other techniques. However, it is unclear whether the dimensions are stable across different samples. Structural consistency within network analysis is defined as how interrelated and homogeneous a dimension is in the presence of other related dimensions (Christensen et al., 2020c). It can be determined by how often a dimension from the original EGA is replicated in the calculated bootstrap EGAs. Christensen and Golino (2021) developed Bootstrap Exploratory Graph Analysis (bootEGA) to assess the reproducibility and generalizability of the results of EGAs. In bootstrap EGA, a sampling distribution of EGA results is generated, until the intended number of repetitions has been created. The outcome of this is a distribution of results relative to which stability can be judged.

A further way of measuring structural consistency is by calculating item stability. This is derived by calculating how often an item is placed in each dimension. This estimates which items specifically contribute to the consistency of the dimensions, and which are inconsistent. There are different reasons an item may be allocated to different dimensions. For example, a small sample size may not provide enough consistency, an item may be multidimensional, that is equally connected to multiple communities. Another cause of unstable item allocation may be item redundancy, where items form a subfactor due to similar semantics. Item stability value of 0.65-0.75 is expected to indicate stability and values below this indicates the item increases instability (Christensen & Golino, 2021). All items with loadings below .75 were dropped. Considering the heterogeneous sample in this study (which is likely to decrease item stability), evaluating item stability is appropriate. In summary, bootEGA allows us to calculate the stability of dimensionality as estimated by EGA, along with item stability.

A secondary statistic computed were network loadings, which are the psychometric equivalence of factor loadings. For each item the node strength for each dimension is calculated. A network loading has been conceptualised as a "node's contribution to the emergence of a coherent sub-network or network" (Christensen et al., 2020c p. 9). To ensure items which increase the stability of the dimensions are retained, network loadings were calculated. Items with the lowest average network loading were dropped and it was run again until all loadings were greater than **.**109. The suggested guidelines for small, moderate, and large network loadings are 0.15, 0.25, and 0.35 respectively (Christensen & Golino, 2021).

Hierarchical Structure and Fit to Data

To validate the results from a network perspective with a latent variable perspective, the fit and potential unidimensionality of the model was calculated using a latent variable perspective. Communities within a network analysis are mathematically equivalent to factors in a latent variable perspective (Golino & Epskamp, 2017) making such direct comparisons

possible. Also, due to the established nature of latent variable models, whether the data can be estimated with bifactor models and further, unidimensionality of the data can be determined.

Therefore, the fit of the model developed through EGA was assessed by a confirmatory factor analysis using lavaan (Rosseel et al., 2022). In order to evaluate the fit, the model was assessed against the Tucker-Lewis index (TLI) and comparative fit indices (CFI) of greater than 0.95 and Root Mean Square Error of Approximation (RMSEA) of lower than 0.08. If the model did not indicate good fit, this suggests the empirical data does not fit this model.

The internal consistency of the scale was assessed using McDonald's omega. Both omega and Cronbach's alpha estimate internal consistency. The Alpha coefficient relies on stringent assumptions being met. For example, alpha requires that the sample is normally distributed and that each item equally measures the underlying construct of the scale while omega does not rely on these assumptions (Rodriquez et al., 2015). Despite the widespread use of alpha, in this case omega is more appropriate.

A bifactor model was estimated. Bifactor models indicate the variance in a scale which can be determined by a general factor that is shared among all items and also specific factors which share further variance compared to the general factor (Rodriquez et al., 2015). This can be estimated using McDonald's omega coefficient. Omega coefficient can also estimate the proportion of the reliable variance that can be attributed to both general factor and the specific factors of a scale. A scale can be clearly unidimensional or multidimensional but can also exhibit aspects of both. Using different models, to evaluate the data from different perspectives, will allow us to consider the most appropriate way of understanding the data.

To estimate unidimensionality, Omega coefficients and explained common variance (ECV) were calculated. The OMEGA function was used to analyse the final structure as a bifactor model within the psych R package (Revelle, 2021). A further way of estimating

unidimensionality is explained common variance (ECV) estimates the extent to which the general factor variance represents the reliable variance of a scale. The higher the ECV, the more likely it is that a scale is unidimensional. The Omega coefficients and ECV are then synthesised to determine how unidimensional the scale is likely to be.

Inter-Scale Correlations To further develop an understanding of the relationship between the MMM and the HADS-D, the presence and search subscales of the MLQ, and the WAS subscales the association between these were evaluated. Spearman's correlation coefficient was calculated due to the ordinal nature of the scales. Of the sample, data was available for 299 participants for the HADS-D, 658 for the presence subscale of the MLQ, and 716 for the search subscale of the MLQ, and for 239 participants for all WAS subscales except the Justice subscale, which had data available for 191 participants.

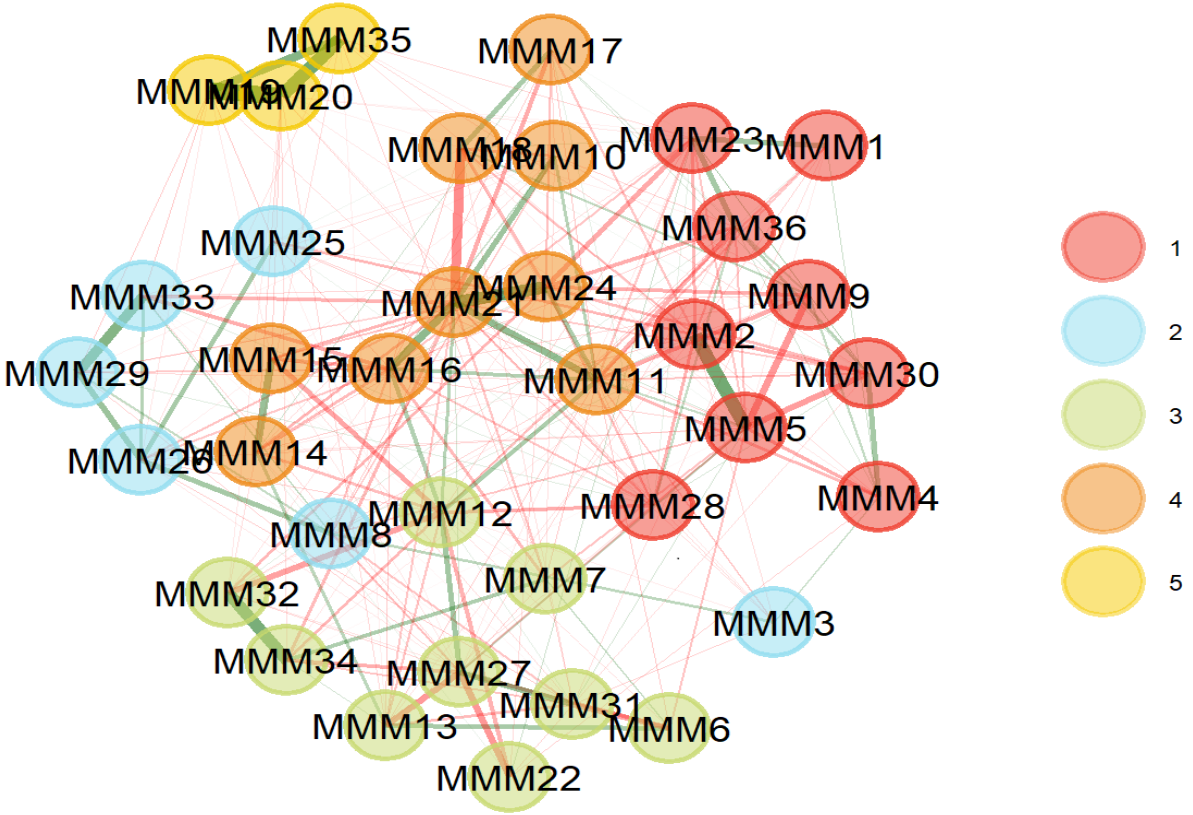
Results

Redundancy

A riEGA using Louvain as a community detection algorithm was conducted in the full sample (N=893). As can be seen in Figure 1, five dimensions were determined. Certain items (such as 20 and 35) appear highly correlated which suggests possible redundancies. To assess item stability, a bootEGA with 500 repetitions was also conducted in the full sample. Item stability was poor (0.59) for one item (22). Therefore, redundancies needed to be addressed.

Figure 1

riEGA conducted in full sample



UVA was completed with all items in the sample (N = 893). Items 19, 20, and 35 were found to be redundant with each other and these three items also load onto their own factor. Out of the three items, 20 is "I often find myself at a loss for what to do next" and appears similar in meaning to "I have trouble feeling a part of my everyday roles" (19) and "I'm not sure how the parts of my life fit together" (35). As it uses easily understood language and exhibits the highest network loading 20 was selected.

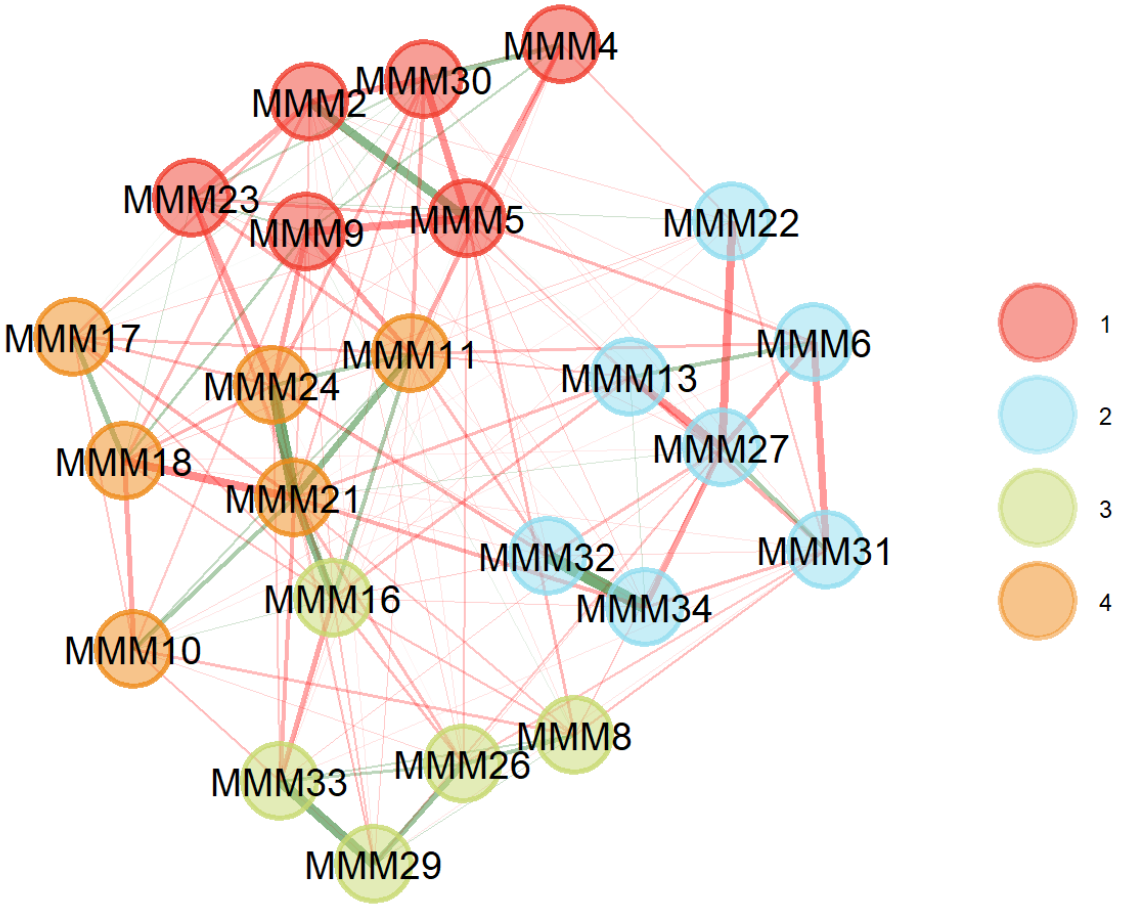
Dimensionality

As can be seen in Figure 2, a visual representation of the final structure suggested a four-factor structure consisting of 24 items. The 24-item sample was analysed using riEGA with Louvain as a community detection algorithm. To obtain a stable structure, riEGA was used to obtain average network loadings for each item in the sample. The item with the lowest

average network loading was removed. The analysis was repeated after removing the item. This was repeated until the loading of each item was greater than .108. As 19 and 35 were already removed due to redundancy, during this process items 20, 1, 3, 7, 12, 14, 15, 25, 28 and 36 were removed. 24 of the original 36 items remained.

Figure 2

Final riEGA 24 item sample



In Table 1, the network loadings for the suggested factors are presented. To interpret these network loadings, small loadings are approximately 0.15, moderate loadings 0.25, and large loadings correspond to 0.35 (Christensen & Golino, 2021).

Table 1

Network loadings for each item included in the final scale

	P*	IDE*	INT*	COH*
2: Other people seem surer than me of where they are going in life.	-0.348	0.062	-0.022	0.061
5: I'm not certain that my life will amount to anything.	-0.242	-0.09	-0.083	0.028
30: I have a strong sense of purpose.	0.234	0.018	0.014	-0.108
9: I feel that my life is going somewhere.	0.169	0.00	0.026	-0.165
4: I feel like I have a mission in life	0.154	0.023	-0.005	-0.044
23: I see a clear path forward for myself into the future.	0.136	0.031	0.038	-0.120
27: I need to stop and think before doing even ordinary things.	0.023	0.363	0.095	-0.013
34: I have a definite idea of my day-to-day priorities.	-0.028	-0.238	-0.016	0.055
22: I don't have to think very hard about what I need to do from moment to moment.	-0.052	-0.238	-0.012	0.032
31: I can't trust my instincts in everyday matters.	0.010	0.226	0.072	-0.005
13: I have confidence when dealing with everyday matters.	-0.040	-0.188	-0.036	0.065
32: Once I get up in the morning, I already have an idea of what I intend to do that day.	-0.020	0.173	-0.021	0.077
6: I can trust my intuition about how to handle ordinary affairs.	-0.047	0.169	0.00	0.031
29: I have come to terms with events that have happened to me in my life.	0.004	0.027	0.237	-0.055
33: I have been able to put the past behind me and move on in my daily life.	0.015	0.025	0.235	-0.080
26: I have been able to make sense of difficulties that I have experienced in my life.	0.040	0.038	0.204	-0.046
16: I've lost the "thread" that used to run through my life.	-0.064	0.076	-0.162	0.188
8: I have been able to find benefit from even my negative experiences.	0.040	0.058	0.135	-0.054
21: Somehow my life has gone off	-0.064	-0.102	-0.227	0.345

18: The story of my life is unfolding in a satisfying way.	0.101	0.009	0.031	-0.229
24: I feel like I am in limbo.	-0.153	0.076	-0.068	0.215
17: I am living up to my potential.	0.059	0.00	0.03	-0.176
11: I have lost sight of my goals.	-0.185	0.11	-0.087	0.173
10: The life I am now leading is not the one I was meant to lead.	-0.020	0.013	-0.097	0.130

P = purpose, IDE = high-level action identification, INT = integration of circumstances, COH = coherence of self-narrative

When the MMM was developed, indicative items for each domain were presented (Brown et al., 2008). Several of the indicative items were omitted from the final sample due to redundancy or lack of stability. Items which were included in the development of the MMM as indicative items and were also present in the final 24-item sample, were used to validate or confirm the meaning of each dimension.

The first dimension consisted of six items, of which two were inversely related to the dimension. This dimension was named ‘purpose.’ Two items which were presented as indicative of purpose in the development of the MMM were placed in this domain in the current structure; “I’m not certain that my life will amount to anything” (5) and “I see a clear path forward for myself into the future” (23). The use of words such as purpose, clear path, and mission all relate to purpose as determined in the development of the MMM (Brown et al., 2008). However, “I have a definite idea of my day-to-day priorities” (34) was expected to be indicative of purpose in the development of the MMM but was instead placed in another dimension (high level action identification). It is likely that this dimension relates to the overarching sense of purpose but does not relate to how a sense of purpose affects our everyday activities.

The MMM was theorised to include one such domain which relates to the relationship between meaning and the experience of routine activities, namely high-level action identification. The item relating to day-to-day priorities (34) was instead placed in the domain

which appears to represent high level action identification. Two items which were hypothesised to be indicative of high-level action identification were “I don’t have to think very hard about what I need to do from moment to moment” (22) and “once I get up in the morning, I already have an idea of what I intend to do that day” (32) were both placed in this dimension. Further, the items included words such ordinary affairs, everyday matters and day to day priorities which suggests this dimension relates to how meaning relates to our daily lives.

A further domain suggested was integration of circumstance. Three items which were hypothesised to be indicative of integration of circumstance were placed in the same domain. These were “I have been able to make sense of difficulties that I have experienced in my life” (26), “I have come to terms with events that have happened to me in my life” (29) and “I have been able to put the past behind me and move on in my daily life” (33). These items relate to being able to make sense of one’s experiences, and they all relate to the experience of challenging circumstances. This suggests that this domain reflects the hypothesised domain of integration of circumstance. On the other hand, the item “I’ve lost the “thread” that used to run through my life” (16) was not originally hypothesized to be related to integration but was grouped within integration of circumstance. It is possible that this item relates to the loss of coherence that is experienced due to a lack of integration. This domain is likely to play an important part in the MMM due its development among people with experiences of trauma.

The last theoretical dimension was coherence of self-narrative. One item which was indicative of coherence of self-narrative was placed in the fourth domain, “the story of my life is unfolding in a satisfying way” (18). While integration of circumstance relates to the ability to make sense of challenging experiences, coherence of self-narrative appears to be indicative of making sense of our lives. While the two dimensions are related, they are

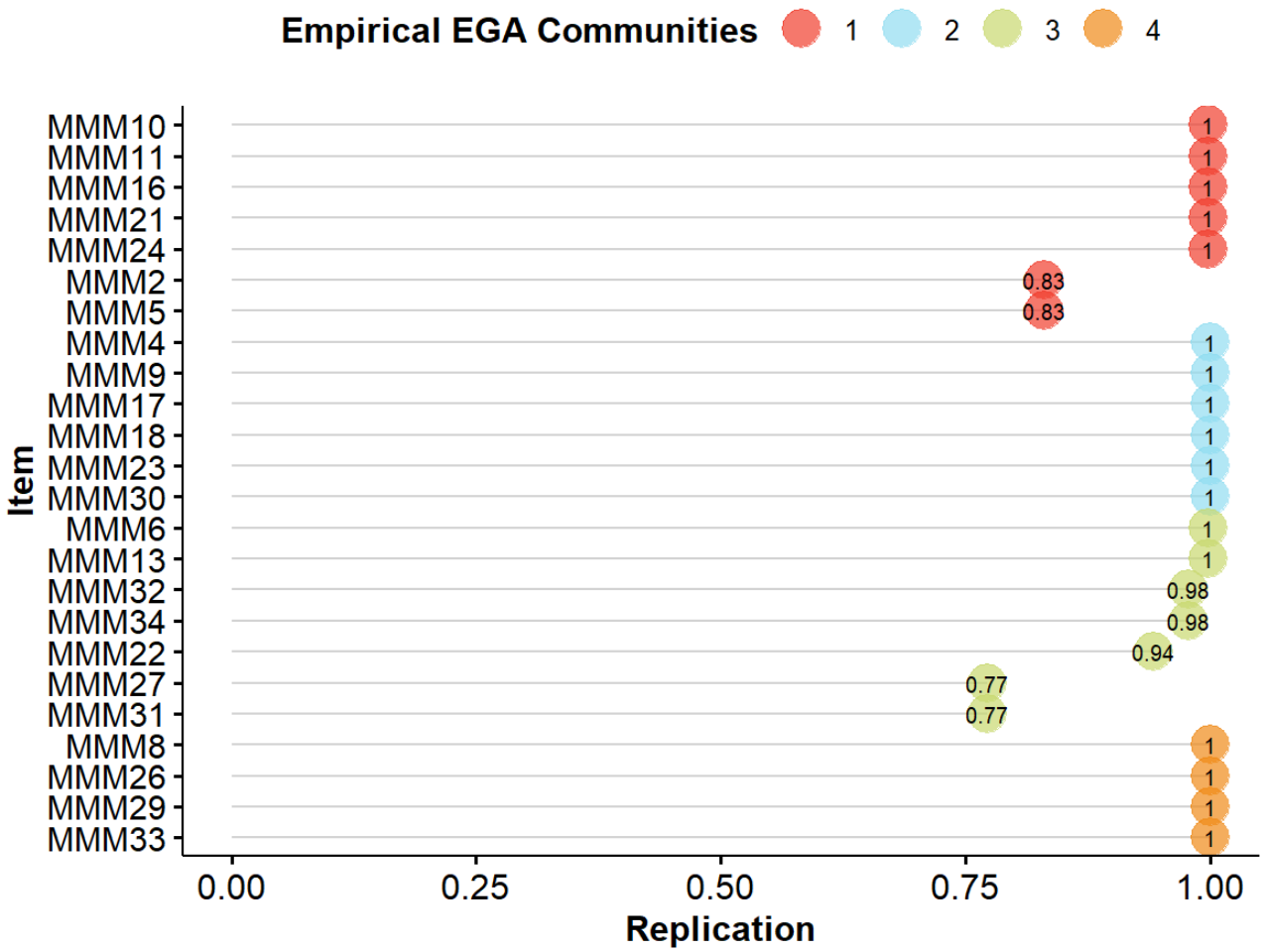
separate facets of meaning in life, which both relate to making sense of ourselves, and our experiences.

Stability

Following this, item stability was assessed. A bootEGA with 500 repetitions was conducted with the 24 items retained in the previous step. The aim was for all items retained to be placed in the same factor at least 75% of the time (Christensen & Golino, 2021). As can be seen in Figure 2, all items showed adequate item stability.

Figure 2

Item stability across 24-item sample



Hierarchical Structure and Fit to Data

A CFA to assess the fit of the model developed using EGA was conducted. This was assessed by the lavaan package in R, and the weighted least squares mean and variance (WLSMV) estimator was used. The factors were allowed to correlate. The model exhibited good fit, $\chi 2$ (246) = 1074.185, $p < 0.001$, CFI = 0.977, TLI = 0.974, RMSEA = 0.061 (95% CI: 0.058-0.065), SRMR = 0.066. As can be seen in Table 2, the factor loadings suggest that the items measure the latent construct. Certain items have negative factor loadings, suggesting that they inversely related to the latent variable. When interpreting the scale, these items need to be reverse scored.

Table 2

Factor loadings for each item included in the final MMM

	item	factor loading	std. error	z-value
Factor 1	MMM4	0.956	0.060	15.886
	MMM30	1.457	0.039	36.905
	MMM5	-1.156	0.057	-20.255
	MMM9	1.439	0.042	34.306
	MMM2	-0.906	0.056	-16.153
	MMM23	1.381	0.043	32.399
Factor 2	MMM22	0.872	0.060	14.595
	MMM6	0.878	0.051	17.079
	MMM31	-0.622	0.073	-8.519
	MMM27	-0.903	0.060	-15.049
	MMM13	1.188	0.050	23.938
	MMM34	1.240	0.045	27.434
	MMM32	1.281	0.047	27.317
Factor 3	MMM16	1.244	0.054	23.105

	MMM33	-1.318	0.048	-27.224
	MMM29	-1.151	0.054	-21.405
	MMM26	-1.125	0.046	-24.262
	MMM8	-1.048	0.051	-20.644
Factor 4	MMM11	1.330	0.049	27.216
	MMM24	1.460	0.048	30.214
	MMM21	1.561	0.043	36.647
	MMM10	0.837	0.068	12.353
	MMM18	-1.511	0.038	-39.494
	MMM17	-1.282	0.046	-27.742

As can be seen in Table 3, an interesting pattern of relationships between the factors emerged. Factor 1(purpose) and factor 2 (high-level action identification) were positively correlated with each other and factor 3 (integration of circumstance) and factor 4 were also positively correlated with each other. However, both factor 1 and factor 2 were negatively correlated with both factor 3 and factor 4. Multiple items are inversely related to factor 3 (except item 16 which relates to the loss of integration of circumstances).

Table 3

Correlations between each factor in the MMM

		covariance	std. err	z-value
Factor 1				
	Factor 2	0.741	0.025	29.304
	Factor 3	-0.757	0.023	-32.405
	Factor 4	-0.861	0.016	-55.375
Factor 2				
	Factor 3	-0.759	0.029	-26.597
	Factor 4	-0.711	0.025	-28.701

Factor 3				
	Factor 4	0.792	0.021	38.637

As can be seen in Table 4, the omega coefficient for the general factor was 0.96, indicating that the 96% of the variance was attributed to a combination of general and specific factors. The hierarchical omega coefficient was 0.75, which suggests that 21% of the reliable variance in total scores are due to the multidimensionality caused by the specific factors. The explained common variance was 0.59, suggesting that 59% of the variance can be explained by the general factor. Also, a comparison between the fit of the model with the general factor combined with the specific factors, $\chi 2$ (186) = 1113.81, $p < 0.001$, RMSEA = 0.075, BIC = -149.99, RMR = 0.04, TLI = 0.8839 compared to just the general factor, $\chi 2$ (252) = 4608.97 $p < 0.001$, RMSEA = 0.139, BIC = -2896.74, RMR = 0.13, TLI = 0.6509 suggested that the combination of general and specific factors was superior. This suggests inconclusive results regarding unidimensionality. On one hand, most of the reliable variance in total scores are caused by the general factor. However, a not insignificant proportion of variance is due to the specific factors and removing the specific factors causes poorer fit of the model. It appears that the scale exhibits both unidimensional and multidimensional properties, but that disregarding the multidimensional aspects of the scale would cause a loss of information that it carries

Table 4

Omega coefficient for each dimension of the MMM

	omega	omega H
General Factor	0.96	0.75
Factor 1	0.85	0.44

Factor 2	0.87	0.62
Factor 3	0.82	0.57
Factor 4	0.88	0.57

Overall explained common variance = 0.59

Inter-scale Correlations

Table 5 presents the correlations between the MMM and its subscales and the HADS-D, a measure of depression, calculated using Spearman's correlation coefficient. The MMM and its four subscales exhibited a moderate, negative correlation with the HADS-D. This suggests that symptoms of depression are inversely related to the experience of meaning in life.

Further, the association between the MMM and its subscales and the presence and search subscale of the MLQ were calculated using Spearman's correlation coefficient. The MMM and the purpose subscale of the MMM were both strongly correlated with the presence subscale of the MLQ. High-level action identification, integration of circumstances, and coherence of self-narrative were moderately correlated with the presence subscale of the MLQ. This suggests that scores on the MMM and the MLQ are positively associated. On the other hand, the relationship between the MMM and its subscales were inversely related to the search subscale of the MLQ. These associations were all weak.

Moreover, the association between the MMM and its subscales and the subscales of the WAS were calculated using Spearman's correlation coefficient. A positive, moderate relationship was found between the MMM and benevolence of the world, self-worth, and luck. A weak relationship was found between the MMM and benevolence of people, justice, control, and also self-control. A weak, negative relationship was found between the MMM and the randomness subscale.

Table 5

Correlation of the MMM with the HADS, MLQ and WAS

	MMM	MMM-P	MMM-IDE	MMM-INT	MMM-COH
HADS-D (N=299)	-0.697	-0.553	-0.603	-0.612	-0.685
MLQ-P (N=658)	0.721	0.749	0.506	0.562	0.642
MLQ-S (N=716)	-0.439	-0.299	-0.374	-0.342	-0.478
WAS-BW (N=239)	0.504	0.464	0.482	0.451	0.451
WAS-BP (N=239)	0.429	0.414	0.436	0.374	0.339
WAS-J (N=191)	0.194	0.184	0.054	0.197	0.192
WAS-C (N=239)	0.097	0.081	0.044	0.109	0.075
WAS-R (N=239)	-0.272	-0.268	-0.192	-0.172	-0.233
WAS-SW (N=239)	0.688	0.643	0.686	0.534	0.6309
WAS-SC (N=239)	0.291	0.294	0.266	0.215	0.234
WAS-L (N=239)	0.573	0.488	0.452	0.550	0.562

Discussion

The aim of this study was to assess the underlying structure of the Measure of Mundane Meaning, which is a measure of meaning in life developed from the perspective of individuals who have undergone experiences which challenge fundamental beliefs and assumptions about meaning, in a general population sample. This was conducted using a novel way of assessing dimensionality (EGA), which does not rely on assumptions of latent variables underlying the data. Evaluating the dimensionality of the MMM lends itself to two separate aims. Firstly, to evaluate the validity of the MMM to assess whether it measures meaning in life as conceptualised during its development (Brown et al., 2008). Secondly, to use the MMM as a proxy of the hypothetical construct of meaning in life, which allows us to broaden our understanding of meaning in life and contrast it with conceptually similar, yet distinct, constructs.

The results of the statistical analysis revealed that, after assessing items for redundancy, poor loadings onto clusters, and inadequate item stability, the final scale consisted of 24 items. The removal of redundant and unstable items has increased the reliability of the questionnaire, as it now captures the underlying construct more precisely. Also, reducing the number of items increased the efficiency and feasibility of the use of the questionnaire. In addition, results suggested the scale was more likely to reflect a multidimensional structure, as opposed to a unidimensional one. This indicates that the MMM measures different facets of meaning in life, which provides a more comprehensive understanding of the underlying construct. Also, a multidimensional structure is in line with previous theoretical structures of meaning in life (King & Hicks, 2021). Coherence between the multi-dimensionality of the theoretical and the actual structure strengthened the validity of the questionnaire.

The structure of the MMM was found to reflect four separate dimensions. One of the advantages of EGA is that items are assigned to each dimension, However, determining the psychological meaning of the underlying dimensions was still necessary. Considering the theoretical rationale for the underlying dimensions provided a framework to evaluate the dimensionality of the MMM. When the MMM was developed, it was theorised to encompass four dimensions: coherence of self-narrative, integration of circumstances, high–level action identification, and sense of purpose (Brown et al., 2008). Indicative items for each dimension were presented. When evaluating the dimensionality of the MMM, many of the items which were included as examples of the hypothesised structure were validated in this study. This suggests a validation of the rationally derived domains of the MMM. The aim of this study was to validate the structure of the MMM to enable a measure of meaning in life which corresponds to its theoretical construct. The current study found sense of purpose, high-level action identification, integration of circumstance, and coherence of self-narrative all formed distinct dimensions, which suggests validity of the questionnaire.

As the four expected dimensions were replicated through this analysis, this indicates that they influence the structure of the MMM. As noted in the introduction, meaning in life has been suggested by others to be made up of only significance, purpose and coherence (King & Hicks, 2021). The difference between the structure of the MMM, compared to this conceptualisation of meaning in life may be understood by the fact that the MMM was developed among people who have experienced a traumatic event. Considering the experiences of someone who is used to experiencing a high level of meaning and then loses that meaning due to trauma, may lead to a different way of understanding and conceptualising meaning. Trauma often results in a loss of meaning, which can explain the difference between the structure of the MMM compared to other conceptualisations of meaning. Further research could examine temporal patterns of responses, as it is possible that

experiences of meaning are more intense following a traumatic event and as these experiences are emotionally processed (either through time or therapy) these response patterns change.

To examine convergent validity of the MMM, the MMM was contrasted with the meaningfulness subscale of the WAS, The MMM was only weakly correlated with the justice and control subscales, and negatively correlated with the randomness subscales, which have been suggested to relate to meaning (van Bruggen et al., 2018). The lack of a relationship suggests that these subscales do not measure the same underlying construct. A sense of justice and an ability to control our experiences, rather than life being random does not appear to be related to meaning in life as measured by the MMM. On the other hand, the MMM was related to assumptions regarding benevolence of the world, self-worth, and luck. This suggests that meaning in life is related to positive assumptions about oneself and the world around oneself. It may be that these positive assumptions are related to a sense of positivity regarding one's future, which may give a sense of meaning to one's life. Also, research has found that trauma often impacts assumptions of self-worth and of benevolence of the world (van Bruggen et al., 2018). As the MMM was developed among a sample with experiences of trauma, this association strengthens the validity of the MMM as a measure sensitive to meaning in life in the context of experiences of trauma.

To further assess for convergent validity, the scores on the MMM were correlated with the presence subscale of the MLQ, a commonly used measure of meaning in life. The MMM and the purpose subscale of the MMM exhibited a strong, positive, relationship with the presence subscale of the MLQ, while high-level action cassociated with the presence subscale of the MLQ. This relationship suggests that the MMM and the MLQ partially measure the same underlying construct of meaning in life. This has provided further evidence for the validity of the MMM. However, the differences between the two scales reflected the

unique perspective, and development of the scales. The MLQ was developed in an undergraduate sample, while the MMM was developed in a clinical sample of adults with experiences of trauma. This may explain why subscales that are likely to be impacted by experiences of trauma were not strongly correlated with the presence subscale of the MLQ. This suggests that the MMM measures meaning in life, but in accordance with its aim and development, has a slightly different outlook on meaning in life.

Also, the relationship between the MMM and the search subscale of the MLQ was a negative, weak correlation. This is similar to the relationship between findings related to the presence and search subscale of the MLQ (Steger et al., 2006). However, cultural experiences of meaning appear to impact this relationship. One study found that, among a Japanese sample search for meaning was positively correlated with the presence of meaning, whilst among an American sample, there was a negative correlation between the search and presence of meaning in life (Steger et al., 2008). This suggests that among the Japanese sample search for meaning was related to increased meaning, while the opposite relationship existed in an American sample, which may have greater cultural similarities to a British sample. The current study's finding of an inverse relationship between the search for meaning and the MMM provided only limited evidence for the validity of the MMM but is consistent with the pattern seen between the presence and search subscale of the MLQ within a Western sample (Steger et al., 2006).

The result of the dimensionality analysis provides further direction for the use of the MMM among clinical populations. The most commonly used measure of meaning in life (MLQ) has not been developed based on a proposed structure of the term meaning in life, or of the factors which make it up. Instead, it uses the dimensions of "presence" and "search" for meaning in life – which do not give any greater clarity about what the term itself means. (Steger et al., 2006). While it has been validated for use in evaluating the experience of

meaning in life in diverse settings and populations, it lacks an evidence-based understanding of a person's experience of meaning in life as a multidimensional construct. The MMM, on the other hand, can be used to understand specific aspects of meaning in life, providing a more precise picture of what meaning in life entails. The combination of the MLQ and the MMM may be appropriate among diverse samples: as the MLQ can evaluate whether there is a broader concern regarding the presence and/or search for meaning in life and the MMM can identify more specific symptoms which are contributing to a loss of meaning. The information from these questionnaires may then be able to guide clinical interventions. For example, for a person with a history of trauma who exhibits a lack of integration of circumstance, a trauma-focused CBT intervention may be appropriate - where the aim of the intervention is to process and integrate trauma. A narrative approach with the aim of developing a coherent narrative may also be effective. On the other hand, if a person shows diminished action identification, behavioural activation may be more effective, as actions which were previously felt as too challenging may become easier over time, increasing high level action identification. It has been theorised that an increase in high level action identification increases meaning in life. As such, the dimensionality assessment of the MMM offers a way of targeting specific aspects of meaning in life in clinical interventions, with the aim of enhancing meaning in life and facilitating recovery.

This analysis contributes to the understanding of meaning in life, as conceptualised by the MMM. Meaning in life is a complex construct with diffuse edges and lacks a shared definition used within research. From the results of the study, underlying aspects of meaning in life have emerged (integration of circumstance, purpose, high-level action identification and coherence of self-narrative). Considering the MMM as a proxy for meaning in life, this perspective provides further understanding of the relationship between meaning in life and similar constructs, namely depression and recovery. It also provides an avenue to consider

whether the MMM can be used in clinical practice, to provide a direction for a clinical intervention, or as a way of measuring recovery.

This conceptualisation gives us a new way of considering the relationship between meaning in life and depression. Depression is a disorder characterised by diminished interest, low mood, and other cognitive, somatic, and emotional symptoms, which cause significant distress or an impairment in functional abilities (American Psychiatric Association, 2013). While many people who experience depression experience a sense of loss of meaning, "loss of meaning" and "depression" are distinct constructs as meaning in life is a hypothetical construct, while depression is a diagnosable disorder. A diagnosis of depression, as described above, requires a specific symptom set to be met. In this study, for example, depression was operationalised as a person's score on the HADS-D. Meaning in life, on the other hand, is a continuum along which people's experience of meaning in life can be measured. But it lacks a threshold along which a clinically significant level of meaning in life can be determined. The relationship between the HADS-D and the MMM was found to be inverse, which is in line with our understanding of meaning in life and depression. However, the dimensionality evaluation of the MMM, as a proxy of meaning in life, may explain the relationship between meaning in life and depression. One of the dimensions in meaning in life is high-level action identification, the theory that behaviour can be experienced at different levels of abstraction (Vallacher & Wegner, 1987). Lower levels of abstraction relate to the mechanics of engaging in a behaviour while higher level abstraction relates to the meaningfulness of the activity. When an activity is difficult, people are more likely to experience it at lower levels of abstraction. This is because the act of doing the activity itself requires more cognitive processing. Therefore, the focus is on the performance of the physical action itself rather than the higher-level purpose of the action. When a person experiences depression, previously simple activity become more difficult. When this happens, behaviours are more likely to lose

higher-level abstraction, and are therefore experienced as less meaningful. This may clarify why a person may express the experience of "going through the motions" when they are depressed. Due to a lack of pleasure and motivation, simple actions, such as showering, become more challenging. When this happens, the activity loses its connection to its purpose, such as caring for one's personal hygiene, which is what gives a behaviour its broader meaning. This can explain the relationship between meaning in life and depression in the evaluation of behaviours. Evaluating the relationship between the underlying dimensions of the MMM and depression has not provided further clarity on the mechanism of the relationship between depression and meaning. While this is a correlational study which would not have been able to determine this mechanism, no further evidence for this mechanism can be gained by evaluating the relationship between each subscale of the MMM and depression, as the associations were all approximately equal strength. Further studies could attempt to determine this relationship by evaluating causal and temporal patterns.

A further construct that the MMM can help us conceptualise and compare is recovery. Meaning in life is used interchangeably with the concept of recovery in mental health as one way of understanding recovery is as the recovery of meaning (Slade, 2009). The CHIME conceptual framework is another way of understanding recovery based on connectedness, hope, identity, meaning and empowerment (Leamy et al., 2011). The MMM was developed among participants who have experienced loss of meaning. Similarly, recovery is a process of rebuilding meaning in one's life following mental health challenges. Therefore, the MMM may be related to recovery as operationalised through the CHIME framework. Loss is an integral aspect of recovery and meaning in life as conceptualised by the MMM. However, both concepts encompass what one can gain when one goes through experiences which challenge one's assumptions of meaning in life. This experience is not about returning to a previous set of assumptions and beliefs, but about rebuilding and integrating one's

experiences to develop a sense of meaning in life. Developing a further understanding of the similarities of meaning in life and recovery may allow us to understand how recovery-focused mental health interventions can be developed. Further research could consider the similarities and relationship between recovery and meaning in life.

Limitations

One of the weaknesses of the present study was the lack of cross-validation of the final structure. Cross-validation includes splitting the sample into multiple datasets to assess the stability and generalizability of the structure over multiple analysis. Due to the sample size, it was not possible to divide the sample further while continuing to achieve meaningful results. Therefore, this result carries the risk of over-fitting the structure to the current dataset. This suggests it is possible that the structure is not generalizable to further samples. Further research may benefit from obtaining larger samples, in order to create two samples of at least 500 which is the suggested minimum sample size (Golino & Epskamp, 2017).

A further limitation of this study is that it is a secondary analysis of a combination of study populations. This heterogenous sample exhibits diverse characteristics and are made up of subgroups, such as people with physical health concerns, university students with no history of mental health conditions, and people who have experienced challenging circumstances such as a miscarriage or loss of a job. Therefore, the heterogeneity of the sample reduces the specificity of the results. Certain subgroups may exhibit different response patterns. This may be a reason why many items were unstable and removed from the final questionnaire. However, Fabrigar and colleagues (1999) warn against the use of homogenous samples which can lead to erroneous loadings and correlations between factors. As such, while heterogenous samples have limitations, homogeneous samples can also cause problems with the analysis. The large sample made up of different subgroups is more likely to be representative of the general population, which is the target population of the measure.

The items which are stable in this heterogeneous population are more likely to be appropriate to use in different contexts.

Further, the sampling procedure used, convenience sampling, is likely to introduce bias to the analysis. Individuals who are more likely to agree to take part in a research project may exhibit different response patterns compared to the general population. Additionally, the lack of information regarding participants' experiences of trauma restricts the analysis of its impact on the MMM. The questionnaire was developed with people with experiences of trauma, which has an impact on meaning in life and it is possible that response patterns depend on experiences of trauma. Further research could determine whether meaning in life, as conceptualised by the MMM, is impacted by traumatic experiences. This analysis could also consider whether the items exhibit differential item stability within more homogeneous groups.

Conclusion

This purpose of this study was to develop evidence for the validity of the MMM by examining its dimensionality. A further aim was to enhance the understanding of the underlying structure of meaning in life, using the MMM as a proxy of the construct. The MMM was found to exhibit a four-factor structure consisting of the following dimensions: purpose, coherence of self-narrative, integration of circumstances, and high–level action identification. Evaluating the underlying structure also of the MMM allows a better understanding of the relationship between depression and meaning in life through the underlying factor of high-level action identification, which can explain how engaging in activities loses its sense of meaning. Determining the structure of the MMM also allows for clinical use of the measure as knowledge about its structure ensures clarity in the interpretation of the measure. This also allows an understanding of how meaning in life is related to the concept of recovery. On the background

of meaning in life as a ‘fuzzy’ construct, this dimensionality analysis has provided both further evidence for the validity of the MMM, and also furthered our understanding of meaning in life.

Systematic review of Measures of Meaning in life

Abstract

Many measures of meaning in life exist. Many of these scales are of variable quality, as determined by a previous systematic review which assessed measures published prior to 2012 (Brandstatter et al., 2012). The aim of the current systematic review was to evaluate the quality of measures of meaning in life published since the previous systematic review. Databases utilised were Medline and Psycinfo. Studies which evaluated the development of self-report tools of meaning in life were included. Inclusion criteria were English speaking, peer-reviewed articles with adult participants. The Cosmin framework was utilised to assess the content validity, internal structure, and remaining measurement properties of each study (Prinsen et al., 2018). Ten studies were included, of which three were developed in clinical samples. For each study, evidence of inadequate methodological quality in their development existed. Generally, this was due to a lack of piloting or cognitive interviewing in the development of the measures. This systematic review has found that there is currently a lack of measures of meaning in life which can be recommended for further use. Clinicians and researchers should be aware of the shortcomings of current measures in their assessment of meaning in life.

The biomedical model of mental disorders posits emotional problems as "biologically-based brain diseases" (Deacon, 2013, p. 847). This theory suggests that mental illnesses, such as depression, are due to structural or functional abnormalities located in the brain and consequently, if these abnormalities are reversed e.g., using psychopharmacology, the person will recover (Deacon, 2013).

The biomedical model has faced significant criticism. Deacon (2013) notes the lack of biomarkers that identify any mental illnesses, suggesting a lack of evidence for the assumptions of the biomedical model. Also, people who are mentally ill are becoming more disabled and are ill for longer despite the development of biological treatments, which suggests the biomedical model has not helped reduce the burden of mental illness, despite huge monetary investments in research and practice.

Clinical psychology is influenced by the biomedical model in a multiplicity of ways, one of which is how treatments are evaluated. Most research within clinical psychology uses symptom measures based on the DSM-5 as outcome measures. For example, one review of studies of psychotherapy for young people found that out of the 236 trials conducted, over 70% focused solely on symptoms reduction as an outcome measure (Weisz et al., 2005). Welch and colleagues (2013) have challenged this use of diagnostic categories within clinical psychology as the validity of diagnostic categories within the DSM-5 has not been sufficiently determined, and poor methodology and secrecy underpinned the development of the DSM-5.

Some researchers argue that outcomes within clinical psychology are more complex than simply a reduction in symptoms. Kazdin (1999) argues that symptom reduction is not necessary for clinically significant outcomes in research. Many other aspects, such as reduced impairment by symptoms or increased quality of life, can dramatically change even as symptom change is static. Kazdin further argues that "psychotherapy is not about reaching a destination

(eliminating symptoms) as it is about the ride (the process of coping with life)" (Kazdin, 2008, p. 147). He suggests there is something beyond symptoms, an ability to live and cope with life, that is a more meaningful outcome.

As discussed above, using only diagnostic categories to evaluate research has significant limitations. To move away from symptom-based outcomes measures, we need to find a new way to evaluate outcomes of treatments and mental health services. The Division of Clinical Psychology suggests that in addition to treating emotional problems the aim of clinical psychology is "to enhance and promote psychological well-being" (Division of Clinical Psychology [DCP], 2001, p.2). Focusing on psychological wellbeing rather than reducing psychological distress are separate aims, which can encourage integration of aspects of a person's life. An alternative to symptom-based outcome measures is a focus on wellbeing.

To use wellbeing as an outcome measure within clinical psychology an agreed upon definition and a valid way of measuring wellbeing is needed. However, wellbeing is an elusive concept with debate about how to operationalise and measure it as a construct. Also, there are several constructs which fit under the umbrella of wellbeing. The most prevalent constructs are eudemonia, psychological wellbeing, meaning in life and within clinical psychology, personal recovery. Huta and Waterman (2014) have argued different definitions are often used to refer to the same construct, which may have led to contradictory outcomes within research. The lack of specificity also means that these definitions risk becoming so broad as to be uninformative (Huta & Waterman, 2014). Developing operational and conceptual definitions of constructs related to psychological wellbeing will provide a framework for developing outcome measures as an alternative to the biomedical framework. I will now share a brief overview of these concepts before focusing in on their relevance to clinical psychology and the current systematic review.

Aristotle introduced the concept of wellbeing (over 2000 years ago) as ‘eudemonia’ which is a Greek word that has subsequently been translated in different ways, one of which is happiness. However, happiness is more appropriately translated as hedonia. Hedonic wellbeing is defined as pleasure and a lack of distress. Some psychologists have argued that the aim of psychological treatment is to increase hedonic wellbeing. Freud, for example, suggests that “the behavior of men themselves reveals as the purpose and object of their lives…The answer to this can hardly be in doubt: they seek happiness, they want to become happy…” (Freud, 1930, p. 23). Freud argues that happiness is the purpose of one’s life, endorsing a hedonic perspective on wellbeing.

However, it has been argued that this translation does not fully capture the meaning of eudemonia (Alexandrova & Fabian, 2022). Other translations of eudemonia refer to “fulfilling one’s virtuous potentials and living as one was inherently intended to live” (Deci & Ryan, 2008 p. 2). In a systematic review by Huta and Waterman, they note that “growth, authenticity, meaning, and excellence… provide a reasonable idea of what the majority of researchers mean by eudaimonia” (Huta & Waterman, 2014, p.1448). These definitions are broad ways of understanding eudemonia. Compared to hedonia which can be seen as a state experience, as “well-being is not so much an outcome or end state as it is a process of fulfilling” (Deci & Ryan, 2006, p.2).

Some see a eudemonic perspective of wellbeing as more important than the hedonic perspective. Jung, for example, argues for a eudemonic perspective on wellbeing. He notes that “meaninglessness… is therefore equivalent to illness” (Jung et al., 1989, p. 340). Jung equates meaning and good health, arguing that meaning is necessary for one to be well. He further argues that “meaning makes a great many things endurable – perhaps everything” (Jung et al., 1989, p. 340). Finding meaning makes pain and distress bearable, and this suggests we focus

on how to increase meaning rather than reducing distress. One of the most influential ways to operationalise eudemonia has been as Psychological Wellbeing by Ryff (1989).

Ryff (1989) has operationalised the concept of psychological wellbeing and created a measure based on this. She has argued that the field of eudemonia lacked valid ways of assessing psychological wellbeing based on research and consequent theoretical underpinnings (Ryff, 1989). Also, she noted the multiplicity of definitions of eudemonic wellbeing which lack specificity and clarity (Ryff, 1989). Due to the gap she identified, she developed one of the primary and most influential measures of psychological wellbeing (Brandel et al., 2017). The six dimensions she argues makes up the concept of wellbeing includes personal growth, self-acceptance, autonomy, environmental mastery, positive relationship, and purpose (Ryff, 1989).

Many further measures of eudemonic wellbeing have been developed, as presented by Brandel and colleagues (2017). They present twelve questionnaires that have been developed to measure eudemonic wellbeing. It is clear there are still many ways of assessing eudemonic wellbeing, with a lack of agreement regarding the most appropriate scales in clinical populations. Only four of the measures in the systematic review (The General Causality Orientations Scale [Deci & Ryan, 1985], Flourishing Scale [Diener et al., 2010], The Mental Health Continuum [Keyes, 2002], The Orientations to Happiness Subscales [Peterson et al., 2005]) have been evaluated in clinical populations (Brandel et al., 2017).

Measures of eudemonic wellbeing have been developed within the field of positive psychology, which focuses on people's strengths and helping people thrive rather than psychological distress. Therefore, measures of wellbeing have been developed in non-clinical populations and are often used in research with 'healthy' participants. Generally, individuals with mental health disorder experience different patterns of wellbeing compared to those with no mental health concerns (Ryff, 2014). Understanding eudemonic wellbeing among people

who have experienced emotional difficulties may require a unique perspective, which is provided through personal recovery.

Within clinical psychology, personal recovery is a concept that takes a eudemonic perspective on wellbeing within the context of emotional problems. Personal recovery focuses on creating meaning within one's illness and living a meaningful life despite one's symptoms; "a way of living a satisfying, hopeful, and contributing life even within the limitations caused by illness" (Anthony, 1993, p. 17). This is similar to Jung's view of the importance of meaning in coping with pain (Jung et al., 1989).

Personal recovery does not approach psychological treatment from a biomedical model with a goal of symptom-free normality since personal recovery was developed within the psychosis field where symptom reduction is not necessarily an achievable goal (Slade et al., 2014). Instead, it allows a more holistic view of a person's ability to live with a mental illness, rather than be defined by it. It requires one to explore what a meaningful life looks like for each individual person (Slade, 2009).

The concept of personal recovery has become an important framework in the development of mental health services in the United Kingdom. Over the last decade mental health policy has transformed mental health services. The governmental strategy No Health Without Mental Health, has been shaped by a focus on personal recovery (Department of Health, 2011). However, when evaluating outcomes within mental health services, the focus on personal recovery has not been translated into practice. Services such as Improving Access to Psychological Therapies (IAPT) still determine recovery based on symptom endorsement and use a mathematical formula to determine 'recovery' (National Collaborating Centre for Mental Health, 2018). The NHS has committed to close to 2 million people being offered mental health support through an IAPT service by 2023/2024 (NHS England, 2019). While the underlying values of the governmental strategy on mental health are based on personal recovery,

determining successful outcomes of the services is still based on the biomedical model. Fully implementing recovery focused psychological support requires organizational transformation (Farkas et al., 2005). Personal recovery has different aims of mental health treatment than clinical recovery, and outcomes measures based on these aims are necessary. Operationalising the concept of personal recovery is necessary to be able to measure the outcome of treatments.

The subjective experience of personal recovery brings up challenges in measurement, as personal recovery looks different among different people. Developing nomothetic measures that truly capture a unique experience is challenging. To assist in the transformation of services clinically valid measures of recovery are needed. Many measures of personal recovery have been developed, with different ways of operationalising personal recovery. However, a systematic review found that no existing measure of personal recovery showed adequate psychometric properties to be recommended above others (Shanks et al., 2013). To determine an appropriate measure of personal recovery, there is a need for a clear definition of personal recovery, with defined boundaries with other concepts.

A framework has been developed to operationalize personal recovery. The connectedness, hope, identity, meaning and empowerment (CHIME) framework is an attempt at developing a conceptual framework of personal recovery (Leamy et al., 2011). Within this framework, connectedness is built on relationships and positive support, hope is characterised as belief and motivation towards recovery and change, identity not based on own's emotional problems and stigma, with meaning as a meaningful life, goals and finding meaning in one's experiences, while empowerment is feeling a sense of strength, control, and responsibility (Leamy et al., 2011). A conceptual framework as presented here suggests an ability to define and operationalise personal recovery. However, the dimensions that define personal recovery are themselves not clearly delineated from other constructs. This is a concern as the challenge in

moving from mental health services based on the biomedical model to eudemonic wellbeing is the lack of conceptual and operational clarity that can provide a basis for measurement.

Concepts such as psychological wellbeing and personal recovery are not clearly delineated constructs despite the development of frameworks to define them. The frameworks that make up psychological well-being and personal recovery are partially based on similar dimensions (Leamy et al., 2011; Ryff, 1989). Dimensions that make up these constructs have not been well-defined. Purpose and meaning are both part of how psychological wellbeing and personal recovery have been operationalised. Also, to further cause confusion, some researchers use the terms meaning and personal recovery interchangeably (Slade, 2009). If meaning is a dimension that is used to define other concepts it is necessary, that meaning itself has a clear conceptual definition. Therefore, meaning is an important part of our understanding of eudemonic wellbeing, and needs a clear and valid definition, or measures that clearly delineate this concept from other similar and related constructs.

A definition of meaning in life (amongst many) suggests that "lives may be experienced as meaningful when they are felt to have significance beyond the trivial or momentary, to have purpose, or to have a coherence that transcends chaos" (King et al., 2006, p. 180). This suggests there are multiple aspects that make up meaning in life, including significance, purpose, and coherence. Leontiev (2013) argues that meaning is complex and that there is no clear definition. He further suggests that meaning is made up of multiple aspects and can be understood in different ways. Leontiev (2013) summarises multiple challenges in the study of meaning, the most relevant for this context being the linguistic, structural, and methodological challenges. The first challenge with research on personal meaning is that there is no definition of the word that is agreed upon within either the academic or common language in English. Leontiev notes that within both German and Russian there are different words that all translate to meaning, suggesting that other languages offer more specific definitions. Meaning is a broad concept in

the English language. Leontiev further suggests that meaning is a multidimensional construct, rather than a single measurable variable. He argues that "meaning is thus a hypothetical construct, something not directly observable but rather conceived. It is needed to explain observable phenomena, but cannot be reduced to them" (Leontiev, 2013, p. 463). Meaning can refer to many different aspects, such as an emotional experience, or both a state or trait variable, and one variable will not be able to include them all (Leontiev, 2013). Leontiev also suggests that one of the major methodological challenges in research on meaning is the lack of discriminant validity. He suggests that our ability to measure meaningfulness will always be impacted by our inability to distinguish meaning from other similar constructs. Meaning is a complex construct which is not clearly defined.

Meaning in life is used to define constructs within eudemonic wellbeing, but itself lacks clarity as a concept. Clearly delineated constructs are necessary to avoid spurious research. Therefore, this work is a systematic review of meaning in life which seeks to give further clarity to the term meaning in life. Despite challenges in defining meaning in life, ways of assessing meaning in life based on different ways of understanding meaning in life have been developed. Most of these assessment measures take the form of self-report questionnaires. Assessing meaning in life will allow us to explore how meaning in life is defined within the research literature.

A previous systematic review of meaning in life assessment instruments (Brandstätter, et al., 2012) found that 59 different assessment instruments had been developed. These assessment instruments measured different aspects such as the presence or search of meaning in life. Few instruments had been validated in languages other than English and the quality of the studies was variable. This systematic review is a similar review to aprevious review published in 2012 (Brandstätter et al., 2012). It is expected that further assessment instruments will have been developed in this time frame. This will allow us to examine how meaning in life

is measured in terms of the standard categories of psychometric validity evidence, including content validity, response processes, structure, and relationship with other variables. The consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidelines for systematic reviews of measurement properties of patient-reported outcome measures were utilised to guide the development of this systematic review (Prinsen et al., 2018).

In summary, the objective of this systematic review is to critically assess the psychometric quality of measures of meaning in life published since 2012. We aim to determine whether there is evidence to suggest instruments of meaning in life exhibit appropriate validity and reliability to be used as outcome measures in clinical psychology This review may address the shortcomings that currently exist due to the lack of clarity regarding meaning in life as a construct, which currently is a significant limitation in the evidence base.

Method

The design of the present review was based on recommendations made by the JBI Manual for Evidence Synthesis (Aromataris & Munn, 2020) and the COSMIN guidelines (Prinsen et al., 2018). The JBI Manual for Evidence Synthesis recommends that the COSMIN guidelines are used as a quality assessment tool for systematic reviews on measurement properties. There are three parts of the COSMIN guideline recommendations in performing systematic reviews of outcome measurement instruments (Prinsen et al., 2018). These are a) guidelines on performing the literature search, b) evaluate the measurement properties, and c) select a patient-reported outcome measure. In this study, only the recommendations on how to evaluate the measurement properties were utilised. This was because of the specific aims of the current study, and therefore COSMIN guidelines were used to evaluate the quality of the development and to review the measurement properties of the assessment tools (Prinsen et al., 2018). This was appropriate due to the aim of the current study, which was to evaluate the

psychometric quality of meaning in life measures However, the COSMIN guidelines on performing a literature search and formulating recommendations were not used (Prinsen et al., 2018). This was due to a low number of measures in the study were developed in clinical populations. As certain measures of meaning in life that are classified as "patient reported outcome measures" as this study aimed to include measures which are appropriate for use in a general population.Instead, to ensure a consistent approach, decisions around the search terms in the literature search were made in line with the previous systematic review (Brandstätter et al., 2012). Also, a recommendation was not able to be followed, namely utilising at least two independent reviewers at each stage of the process. This was not possible due to the lack of resources available for this project.

Identification and selection of studies

A search for articles was conducted in the electronic databases psycinfo and MEDLINE and the citations were extracted, and duplicates removed. Following this, the titles and abstracts of all articles identified in the search were screened by one reviewer to determine whether they appeared to satisfy the inclusion criteria. Next, the full texts of these articles were retrieved and reviewed against the inclusion and exclusion criteria by the same reviewer.

Search Terms

The search terms used to identify the construct of meaning in life were 'meaning in life' and 'purpose in life.' Since the present review was an update on a previous systematic review, these search terms were determined based on Brandstätter's (2012) review. Boolean operators were used with the following search terms: ("meaning in life" AND/OR "purpose in life") AND (scale* or test* or questionnaire* or assessment* or measur* or inventor* or instrument*).

'Purpose in life' has been included as a search term as it has been used to refer to meaning in life in multiple contexts (Hill et al., 2015). An initial questionnaire developed and

validated to measure meaning and purpose in life was named the Purpose in Life test (Crumbaugh, 1968). Since, in research, the two terms have been conflated and are often used to refer to the same construct. However, this is problematic as people understand meaning and purpose from a layman's perspective as slightly different constructs (Hill et al., 2015). Based on these considerations, including purpose in life as a search term would enable us to capture the full breath of assessments measures developed. However, if a study specifically noted that they were not evaluating meaning in life, these were not included even if they referred to 'purpose in life.'

Criteria for studies chosen for the review

Inclusion criteria

Articles that describe the development of self-report tools evaluating meaning in life were included. The search included articles published between January 2011 until January 2022. Measurement instruments published before 2011 were included in the prior systematic review (Brandstätter et al, 2012) and have already been assessed. Assessment tools developed using both clinical and non-clinical samples were included as non-clinical samples are appropriate in the initial development of a construct.

Exclusion criteria

Exclusion criteria were studies which assessed meaning in life among individuals under the age of 18, those published in languages other than English, and studies that were not peer-reviewed. Articles where the instruments were used as an outcome measure were excluded as the goal of these articles is unrelated to the psychometric properties of the outcome measure. This approach is recommended by the JBI chapter 12.2.5 (Aromataris & Munn, 2020).

Data extraction

For each study, as per the COSMIN guidelines, the intended construct, target population, mode of administration, subscales, number of items, response options, range of

scores, original language, and language of available translations were extracted (Prinsen et al., 2018). Following this the characteristics of the study population were extracted (Prinsen et al., 2018). The number of participants, as well as their age, gender, and ethnicity were noted. Also, for clinical populations the type of disease, its duration and severity were extracted. Further, the setting of the study, the country and language it was conducted in, and the response rate was noted.

Assessment of Methodological Quality

A psychometric quality assessment took place using COSMIN's 'criteria for good measurement properties.' There are three stages to assess measurement properties: evaluating content validity, internal structure, and remaining measurement properties.

There are three steps to evaluating content validity. Firstly, the quality of the development of the measure is rated. Following this, the quality of any content validity study is evaluated. Lastly, specific criteria for the relevance, comprehensiveness and comprehensibility of the questionnaires are assessed using questions such as whether an appropriate qualitative data collection method was used in order to identify items (Prinsen et al., 2018). A second reviewer also rated half of the studies to ensure reliability of the ratings, evaluating the scale items for relevance, comprehensiveness, and comprehensibility. Agreement between the reviewers was calculated to ensure accuracy of the ratings.

If there is high quality evidence for insufficient content validity, it is recommended that further evaluation of the instrument is not conducted (Prinsen et al., 2018). All criteria are made up of specific standards which lead to a rating of very good, adequate, doubtful, inadequate, and not applicable. For each section the 'worst score counts' principle is applied, and an overall rating is given. The overall score was then rated as sufficient (+), insufficient (-), indeterminate (?) and inconsistent (±) based on the COSMIN criteria.

Following this, the internal structure of the scale was evaluated by examining structural validity, internal consistency and cross-cultural validity. Thereafter the remaining measurement properties such as reliability, measurement error, criterion validity, hypotheses testing for construct validity and responsiveness were assessed, if available. The score for these measurement properties were rated as sufficient (+), insufficient (-) or indeterminate (?). To enhance the reliability of the methodological assessment a second reviewer conducted an independent methodological assessment using the same criteria as the author to assess five randomly selected studies. Agreement between the reviewers was calculated to assess the reliability of the ratings.

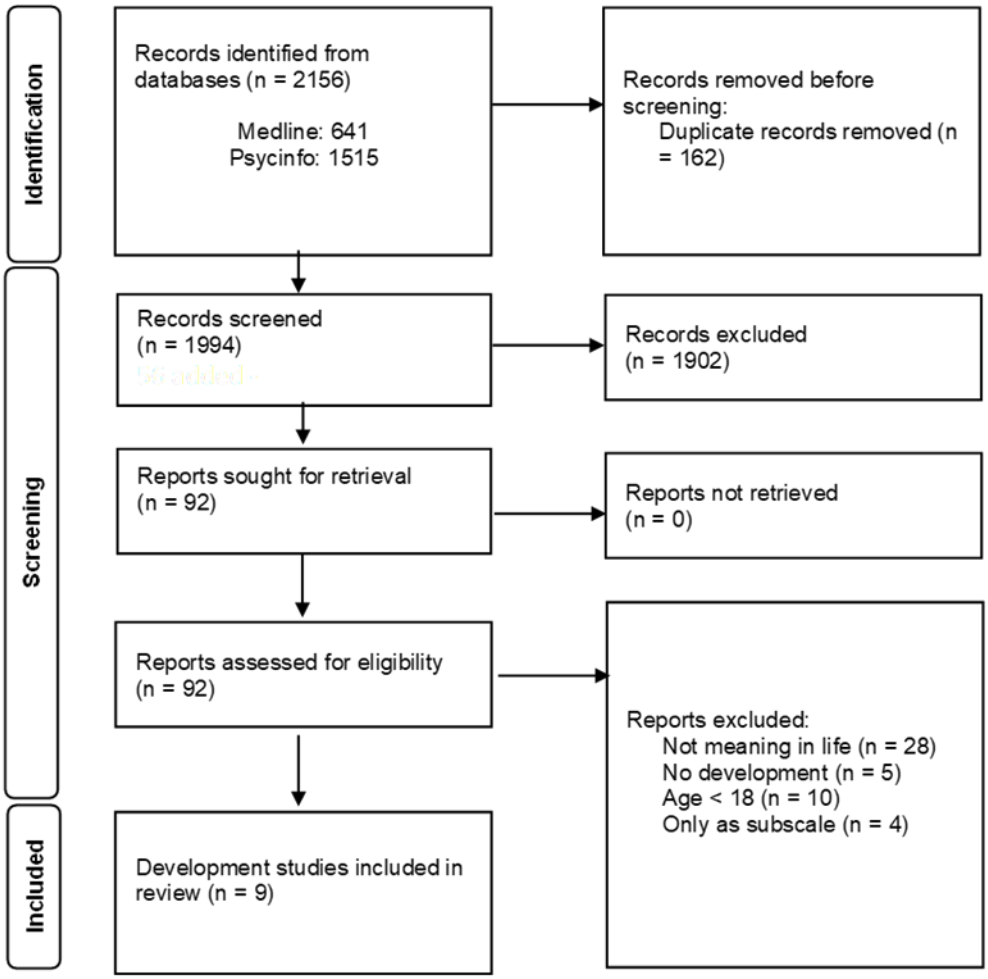
Next, the methodological quality of the studies was assessed using the COSMIN Risk of Bias checklist (Prinsen et al., 2018). Following this, the evidence was summarised and graded using a modified Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach as presented in the COSMIN guidelines (Prinsen et al., 2018). This means evidence can be rated as high, moderate, low, or very low level quality of evidence. A narrative synthesis was then developed to provide a coherent narrative of the results.

Results

Based on the initial search strategy 2156 articles were found, with 641 from Medline and 1515 from Psycinfo (see Figure 1). After removing 164 duplicate articles 1994 records remained. These were screened based on titles and/or abstracts and 1902 were excluded at this stage, leaving 92 records. Full text articles were retrieved for the remaining 92 records. A further 82 were excluded based on the exclusion and inclusion criteria. Ten articles were included in the present review.

Figure 1

Flowchart of study selection



Article characteristics

Of these ten questionnaires, four scales measured meaning in life (Wang & Liao, 2015; George & Park, 2017; Xia et al., 2017; Guerra et al., 2017). Two scales measured meaning and purpose in life (Salsman et al., 2020; Schulenberg et al., 2011). One study examined meaning in life and sources of meaning in life (Zhou et al., 2021), and another study examined HIV-specific meaningfulness of life (Audet et al., 2015). A further study included need for meaning, meaning confusion, meaning avoidance, and meaning anxiety (Zhang et al., 2018) while the

last scale evaluated meaning in life as felt sense, mattering/significance, purpose/goals, coherence, and reflectivity (Hill et al., 2019).

As can be seen in Figure 2, each scale was a self-report questionnaire made up of between 4 and 37 items. Scales that were not unidimensional, were made up of between two and six subscales, and were rated separately. The response options for each scale were Likert-style scales consisting of between five and nine points. Five scales were developed in English, four in Chinese and one in Portuguese. The studies were developed in the USA, China, Taiwan, and Portugal.

Participant characteristics

The ten scales were developed in 18 separate samples. Of the samples five had an average age of between 18-25, two between 26-40, nine between 41-59 and one above 60. Data for one sample was missing. Two samples were made up of either only female or male participants, with the remaining samples consisting of between 27% and 74% who identified as either female or women. Data for one sample was missing. Ethnicity was only provided in studies in American settings, where the majority of participants were white. In one study, the majority of participants identified as African American, which was in the study aimed to determine a way of measuring meaningfulness in life in an HIV positive sample (Audet et al., 2015). The studies conducted in China and Portugal did not offer information on the ethnic background of the participants.

Seven of the scales were developed in general population samples, such as among university students, or through online research panels. One of these was specifically developed among the elderly population. Three of the scales were developed in clinical populations made up of cancer and HIV patients, which took place in hospitals or health clinics.

Clinical Sample Characteristics

Three scales were developed within clinical samples (Audet et al., 2015; Guerra et al., 2017; Xia et al., 2017). The aim of the scale created by Audet and colleagues (2015) was to develop a way of measuring meaning in life among individuals with HIV/AIDS. The average disease duration for participants with HIV was 8.1 years (SD = 5.6) and for AIDS 2.3 years (SD=0.5). An indication of the severity of the disease for the participants was provided by the average CD4 cell count which was 432 (SD = 338). The variability of the results suggests participants were at different stages of the disease progression. However, this information was only available for 75 of the participants.

Guerra and colleagues (2017) developed a scale to measure meaning in life in both healthy and clinical populations. They included participants with colorectal cancer, breast cancer and prostatic cancer or spinal cord lesion as well as a healthy sample. They did not provide any further information on disease duration or severity, so it is not possible to evaluate whether this sample is comprised of individuals at different stages of disease progression.

Xia and colleagues (2017) aimed to develop a Chinese version of a meaning in life scale that is appropriate for use with cancer patients. Participants with breast cancer, lung cancer, gastric cancer, colorectal cancer, gynaecologic cancer, and nasopharyngeal cancer were included in the development of the questionnaire. They did not provide information on disease duration. However, they reported that of the first sample 4 (16%) had stage one cancer, 7 (28%) had stage two cancer, 5 (20%) had stage three cancer, and 9 (36%) had stage four cancer. Of the second sample, 2 (10%) had stage one cancer, 3 (15%) had stage two, 7 (35%) had stage three and 8 (40%) had stage four. Of the last sample 23 (9%) had stage one, 42 (17%) had stage two, 90 (36%) had stage three and 96 (38%) had stage four cancer. The participants at different stages of disease progression indicates a variability of experience which increase external validity.

Table 1.

Characteristics of the included articles

Reference	Construct	Target population	Mode of administration	Subscales Number of items	Response options	Range of scores/scoring	Original language	Available translations
Schulenberg, S. E., Schnetzer, L. W., & Buchanan, E. M. (2011).	Perceived meaning and life purpose	General population	Self-report questionnaire	4 items	7-point Likert scale	4-28	English	Unknown
Audet, C. M., Wagner, L. J., & Wallston, K. A. (2015).	HIV-specific meaningfulness of life	HIV + patients	Self-report questionnaire	4 items	7-point Likert scale	4-28	English	Unknown
Wang, Y.-H., & Liao, H.-C. (2015).	Meaning in life	General population within Taiwanese cultural context	Self-report questionnaire	33 items	5-point Likert scale	33-165	Chinese	English
George, L. S., & Park, C. L. (2017).	Meaning in life	General population	Self-report questionnaire	15 items, 3 subscales of 5 items each	7-point Likert scale	1-7 for each of the 3 subscales	English	Unknown
Guerra, M. P., Lencastre, L.,	Meaning in life	Within Portuguese	Self-report questionnaire	7 items	5-point Likert scale	7-35	Portuguese	English

Silva, E., & Teixeira, P. M. (2017).		medical setting						
Xia, H.-Z., Gao, L., Wang, Y., Song, H., & Shi, B.-X. (2017).	Meaning in life	Cancer patients	Self-report questionnaire	25 items	5-point Likert scale	25-125	Chinese	English
Zhang, H., Sang, Z., Chen, C., Zhu, J., & Deng, W. (2018).	Need for meaning, meaning confusion, meaning avoidance, and meaning anxiety	General population	Self-report questionnaire	4 subscales, 34 items	7-point Likert scale	34-238	Chinese	English
Hill, C. E., Kline, K. V., Miller, M., Marks, E., Pinto-Coelho, K., & Zetzer, H. (2019).	Meaning in life as felt sense, mattering/ significance, purpose/ goals, coherence, and reflectivity	General population	Self-report questionnaire	2 subscales, 8 items	9-point Likert scale	8-72	English	Unknown
Salsman, J. M., Schalet, B. D., Park, C.	Meaning and purpose in life	General population	Self-report questionnaire	37 items 8 items 6 items	5-point Likert scale	37-185 8-40 6-30	English	Unknown

L., George, L., Steger, M. F., Hahn, E. A., Snyder, M. A., & Cella, D. (2020).				4 items		4-20	
Zhou, J.-J., Tong, P., Ren, Q.-Z., Li, T., Zheng, Y.-J., Shen, Q.-Q., Liang, Y.-Y., & Gao, Y.-L. (2021).	Meaning in life and sources of meaning in life	Elderly	Self-report questionnaire	6 subscales, 28 items	7-point Likert scale	28-196	Chinese

Table 2

Characteristics of the included study populations

		Population			Instrument administration			
Reference	N	Age; Mean (SD) in years	Gender % female	Ethnicity %	Setting	Country	Language	Response rate
Schulenberg, S. E., Schnetzer, L. W., & Buchanan, E. M. (2011).	298	M = 19.7, SD = 2.2	63.3%	72% White 21% Black 3% Asian/Pacific Islander 3% Hispanic 2% American Indian/Alaskan native 2% ‘‘Other’’	Medium sized university	USA	English	
Audet, C. M., Wagner, L. J., & Wallston, K. A. (2015).	125	M = 41.7, SD = 2.0	27%	61% African American 39% Caucasian	Care centre for HIV	USA	English	
Wang, Y.-H., & Liao, H.-C. (2015).	500	M = 42.3, SD = 1.85	62%		Taiwanese public, across different demographic areas and backgrounds; at schools, workplaces, or in public spaces	Taiwan	Taiwanese	95.2%
George, L. S., & Park, C. L. (2017).	188 262 160	19 (median) 19 (median) NA	68.6 % 64.9 % NA	77.7 % white 69.5 % white NA	Large university	USA	English	
Guerra, M. P., Lencastre, L.,	200	M = 57, SD = 8.16	49%		Hospital	Portugal	Portuguese	2 missing participants

Silva, E., & Teixeira, P. M. (2017).	150	M = 54.8, SD = 11.24	100%		Hospital			
	92	M = 59, SD = 12.28	0%		Rehabilitation centre			
	88	M = 32.2, SD = 13.57	62.5%		Education institution			
Xia, H.-Z., Gao, L., Wang, Y., Song, H., & Shi, B.-X. (2017).	25	M = 43.6, SD = 15.25	56%		Hospital	China	Chinese	
	20	M = 42.1, SD = 11.97	55%					
	251	M = 44.4, SD = 13.42	54.2%					
Zhang, H., Sang, Z., Chen, C., Zhu, J., & Deng, W. (2018).	1143	M = 24.73, SD = 2.75	42%		University	China	Chinese	1063 after removing poor quality data
Hill, C. E., Kline, K. V., Miller, M., Marks, E.,	473	M = 19.64 SD = 1.49	74%	57% White, 18% Asian/Pacific Islander, 11% Black, 7% Hispanic/Latino, 6% “other”	University	USA	English	
Pinto-Coelho, K., & Zetzer, H. (2019).	401	M = 35.5, SD = 12.14	62%	75% White, 6% Asian/Pacific Islander, 7% Black, 6% Hispanic/Latino, 5% “other”	Mturk (online research panel)			
Salsman, J. M., Schalet, B. D., Park, C. L., George, L., Steger, M. F.,	1000	M = 47.8, SD = 16.2	49.7%	68.3% White 20.0 % Black/African American 4.1% Asian or Pacific Islander 413.7% Native American or	Online research panel	USA	English	

Hahn, E. A., Snyder, M. A., & Cella, D. (2020).				Alaskan Native 371.4% Native Hawaiian or Other Pacific Islander 5.0% 'Other'			
Zhou, J.-J., Tong, P., Ren, Q.-Z., Li, T., Zheng, Y.-J., Shen, Q.-Q., Liang, Y.-Y., & Gao, Y.-L. (2021).	601	M = 77, SD = 9.34	57.2%		Community health centre	China	Chinese

Synthesis of evidence

The overall rating and quality of evidence for the measurement property of each study is presented below in Table 3. The COSMIN guidelines suggest that each subscale is rated individually, but all subscales which received identical scores have been presented together. An instrument made up of two subscales, one of which is reflectivity, received different ratings (Hill et al., 2019) and are therefore presented in two separate columns. As none of the articles included reported measurement error, this was excluded.

Table 3

Measurement properties for each study

Internal consistency	Cross-cultural validity	Reliability	Criterion validity	Construct validity	Responsiveness
+			+	+	
moderate			moderate	moderate	
?		?		+	?
moderate		moderate		moderate	very low
+		?		+	
moderate		moderate		moderate	
?		?		+	
moderate		moderate		moderate	
?				+	
moderate				moderate	
+					
moderate					
?				+	
moderate				moderate	
?		?		+	
moderate		moderate		moderate	
?		?		+	
moderate		moderate		moderate	
+				+	
moderate	moderate			moderate	
?		+			
moderate		moderate			

Reference		Content validity	*Relevance*	*Comprehensiveness*	*Comprehensibility*	Structural validity
Schulenberg et al., 2011	Overall rating	±	?	-	+	+
	Quality of evidence	very low	Very low	very low	very low	moderate
Audet et al., 2015	Overall rating	?	-	?	?	?
	Quality of evidence	very low	Very low	very low	very low	moderate
Wang et al., 2015	Overall rating	?	-	?	?	+
	Quality of evidence	very low	Very low	very low	very low	moderate
George et al., 2017	Overall rating	+	?	+	+	-
	Quality of evidence	very low	Very low	very low	very low	moderate
Guerra, et al., 2017	Overall rating	?	?	-	?	-
	Quality of evidence	very low	Very low	very low	very ow	moderate
Xia et al., 2017	Overall rating	±	±	-	?	+
	Quality of evidence	very low	Very low	very low	very low	moderate
Zhang et al., 2018	Overall rating	?	+	?	?	-
	Quality of evidence	very low	Very low	very low	very low	moderate
Hill et al., 2019 (A)	Overall rating	±	?	-	+	-
	Quality of evidence	very low	Very low	very low	very low	moderate
Hill et al., 2019 (B)	Overall rating	±	?	+	+	-_
	Quality of evidence	very low	Very low	very low	very low	moderate
Salsman et al., 2020	Overall rating	+	+	+	+	+
	Quality of evidence	very low	Very low	very low	very low	moderate
Zhou et al., 2021	Overall rating	?	?	+	?	-_
	Quality of evidence	very low	Very low	very low	very low	moderate

Quality rating

The methodological quality of the development of each of the measures included was rated as 'inadequate' according to the COSMIN criteria, but the quality of this evidence was 'very low.' This suggests none of the instruments were developed with adequate methodological quality and that the instruments themselves are likely of poor quality, but we cannot rely on this evidence due to the poor quality of the evidence itself. COSMIN guidelines state that further assessment of measurement properties should not be conducted if there is high quality evidence of inadequate development methodology (Prinsen et al., 2018). Since the quality of evidence for each study was considered 'very low' further assessment is appropriate, while keeping in mind the inadequate development of each instrument.

Content validity

Content validity was rated for each study, which included relevance, comprehensiveness, and comprehensibility of the items that make up each instrument. For all but two of the studies overall content validity was rated as indeterminate or inconsistent. The quality of evidence for content validity was rated as 'very low' for each study, mainly due to the poor methodology of the development of the measures. One of the subscales by George and colleagues (2017) and Salsman and colleagues (2020) instrument were rated as sufficient overall content validity. Only Salsman and colleagues (2020) measure was rated as sufficient for relevance, comprehensiveness, and comprehensibility, indicating it is the most likely to have sufficient content validity. However, despite this it also exhibited 'very low' quality of evidence.

Structural validity

Each of the ten studies examined the structural validity of the measures. The methodological quality of the evidence for structural validity for of each of the measures was rated as 'moderate' according to the COSMIN criteria. The measures with evidence of

sufficient structural validity were Schulenberg and colleagues (2011), Wang and colleagues (2015), Xia and colleagues (2017) and Salsman and colleagues (2020). This was based on confirmatory factor analyses with CFI (comparative fit index) or TLI (Tucker Lewis Index) or comparable measure greater than 0.95 or RMSEA (Root Mean Square Error of Approximation) of less than 0.08. The other studies exhibited insufficient or indeterminate structural validity, which meant they either did not provide the information or provided information that the confirmatory factor analysis did not meet the criteria above. If a study only reported the results for an exploratory factor analysis the following criteria were used to determine sufficiency: the Kaiser-Meyer-Olkin Test for sampling accuracy of above 0.8 and a significant result for Bartlett's test of sphericity (Kaiser, 1974).

Internal consistency

Internal consistency was reported for each study. Seven of the studies were found to exhibit indeterminate evidence of internal consistency, while four of the studies showed evidence for sufficient internal consistency (Schulenberg et al., 2011; Wang et al., 2015; Xia et al., 2017 and Salsman et al., 2020). For these four studies, the Cronbach's alpha of was .7 or higher for each subscale of the measures. Also, the COSMIN guidelines require evidence for sufficient structural validity to be considered having sufficient internal consistency (Prinsen et al., 2018). The quality of evidence for each study was rated 'moderate.'

Hypothesis testing for construct validity

All studies that examined construct validity fulfilled criteria for sufficient construct validity. Each study was rated as exhibited moderate quality evidence for this conclusion. For each study, over 75% of the hypotheses set by researchers were found to be in accordance with the expected results. Examples of hypotheses were expected positive correlations between the measures of meaning in life and other established well-being measures and negative

correlations with measures of anxiety and depression. Two studies, Xia and colleagues (2017) and Zhou and colleagues (2017) did not examine construct validity.

Criterion validity

One study examined criterion validity. According to the COSMIN criteria, the only 'gold standard' is a longer version of a measure when we are evaluating a newer, shorter version (Prinsen et al., 2018). Schulenberg and colleagues (2011) developed a short version of the Purpose in life test (PIL-SF). The correlation between the PIL-SF and the original measure was above .70 which suggests sufficient evidence for construct validity. The quality of evidence was considered moderate.

Reliability

Six studies examined reliability of the measures. The quality of each study was rated as moderate, however five of the six studies exhibited indeterminate reliability. These studies did not report the intraclass correlation coefficient (ICC) or weighted kappa for reliability. Only Zhou and colleagues' measure (2021) was rated as having 'sufficient' reliability. This measure had an ICC of .856 and a week-long break between first and second administration.

Responsiveness

One study examined the responsiveness of an HIV meaningfulness measure (Audet et al., 2015). The responsiveness of the measure was found to be 'indeterminate,' and the quality of this evidence was rated as 'very low.' The researchers compared scores for participants with HIV/AIDS undergoing an expressive writing intervention. They compared this to control group undergoing a neutral writing intervention and compared their scores before and after the intervention (Audet et al., 2015). Neither group exhibited a significant change in their scores in the two months between baseline to one-month post intervention follow up (Audet et al., 2015). A paired t-test was used to measure, which is an inappropriate statistical method to determine valid change and therefore responsiveness (Prinsen et al., 2018). Neither the

statistical method nor the conclusion of the study suggested the measure is appropriately responsive to change.

Overall quality rating

The development of each instrument rated by the COSMIN criteria was deemed to exhibit inadequate quality. The measure which showed the highest proportion of evidence for sufficient validity was Salsman and colleagues' (2020) measure to assess meaning and purpose. It was the only measure that showed evidence for comprehensibility, comprehensiveness, and relevance of the measure, suggesting sufficient content validity (although this was very low-quality evidence). It also exhibited sufficient evidence for structural validity and internal consistency. As all other instruments that examined construct validity, it exhibited sufficient evidence. It was the only measure to assess cross cultural validity, which was rated as sufficient. All other instruments were developed from a classical test theory perspective, and this instrument was the only one that used item response theory in its development. While it was the measure with the highest level of evidence, it also exhibited inadequate development according to the COSMIN criteria (Prinsen et al., 2018). Therefore, it cannot be recommended for widespread clinical use.

To assess the inter-rater reliability of the methodological assessment ratings, Cohen's kappa was calculated for the five studies which a second reviewer assessed. The kappa value of 0.79 indicated high level of agreement between the two author and the second reviewer. This suggests higher confidence in the methodological assessment ratings.

Discussion

The aim of the present review was to critically assess instruments that measure meaning in life published since 2012. The COSMIN guidelines were used to achieve this aim (Prinsen et al., 2018). This systematic review included ten measures aimed at assessing the presence of meaning or purpose in life and one study which also examined sources of meaning in life (Zhou

et al., 2021). A previous systematic review of measures of meaning in life was published in 2012, and this review provides an update on the evidence base published during the last decade.

Instruments of high quality could offer a way of managing challenges that exist in the experimental and clinical work on eudemonic wellbeing. One way that they could do this is by providing an alternative to symptom-based measures. These instruments could support mental health services in their transition to a person-centred recovery-based approach. However, many measures exist which have vague aims and are of poor quality, so critical assessment is needed to ensure that the most appropriate measures are recommended for clinical use. Further, high quality measures could play a role in defining and providing further clarity on meaning in life as a construct, by using the scales as proxies for the construct. This could help provide evidence to delineate between similar but different eudemonic constructs: such as personal recovery and psychological wellbeing. This is necessary to avoid research outcomes which lack clarity and specificity.

This systematic review did not find evidence of high-quality instruments. Firstly, there was evidence of inadequate methodological quality in the development of each instrument included in this review. Also, there was a lack of clarity regarding the underlying construct of meaning in life and what the instruments aimed to measure. Some instruments included new aspects of meaning in life which have not been considered before and therefore limits the continuity and applicability of previous research to these findings. There was also inconsistent evidence provided on the underlying structure of the instruments. Also, the lack of diversity in the cultural contexts which these instruments were developed in limits the generalizability of these instruments. An additional aspect in evaluating these studies is the reliance on classical test theory in the development and evaluation of the instruments. I will expand on these conclusions below.

The quality of the development of all instruments in this systematic review was rated as inadequate, which impacts not only content validity of the instruments but the conclusions that can be reached regarding all measurement properties. Often, the development of the items that make up the measures was based on a review of the literature and opinions of the researchers, rather than including any input from the target population. This means that it is likely that the content of the items lacks an evidence base to determine whether they are an appropriate reflection of the construct. Only three studies included cognitive interviewing or another type of pilot study in the development of the questionnaires (Xia et al., 2017; Salsman et al., 2020; Zhou et al., 2021). A pilot study is an important part of ensuring the quality of the instruments and is recommended by the COSMIN guidelines (Prinsen et al., 2018). Knafl and colleagues (2007) suggest that cognitive interviewing is a necessary part of evaluating participants' understanding of the items and ensures the quality of an instrument. In the development of these measures, there was a reliance on statistical evidence (such as a factor analysis) to determine and assess the items selected for the final measures. Overall, it is not clear whether the target populations understood the items or felt the items were a relevant and complete representation of meaning in life. This suggests that recommending these instruments for widespread use, whether that is for clinical or experimental use, would not be appropriate. A further concern regarding the inadequate quality of the instruments was reliability, which was generally of poor quality. Only one study had sufficient evidence to suggest the results were reliable (Zhou et al., 2021). The lack of evidence for consistent results is concerning. This suggests these measures are not of adequate quality to be recommended for use in other settings. However, compared to the inadequate development of items, which cannot be modified afterwards without creating a new instrument, reliability is a measurement property that can be evaluated in further studies.

In addition to the inadequate methodology of the development of the instruments, there was a lack of clarity regarding what constructs were assessed. For example, two of the included studies (Schulenberg et al., 2011; Salsman et al., 2020) stated that their aim was to examine both purpose and meaning. The studies did not distinguish between these two concepts, and it can be assumed that the authors used them interchangeably as purpose and meaning have been used in this way in previous research (Hill et al., 2015). However, if uncertainty exists as to whether these words are used to refer to the same concept, it challenges the ability to draw conclusions about an instrument and also about meaning in life. Therefore, it is critical to use consistent terms in the development of an instrument. Otherwise, it becomes difficult for the research on meaning in life to develop an agreed upon definition and produce consistent evidence.

Despite this, all studies (except one; Guerra et al., 2017) scored highly on the COSMIN criteria of whether the definition and origin of the construct was clearly described. While the COSMIN guidelines expect a clear and specific description of the construct, the origin of the construct can be "a theory, conceptual framework or disease model used, or a clear rationale provided to define the construct" (Prinsen et al., 2018). McKenna and Heaney (2021) argue that the COSMIN guidelines do not adequately consider the specificity of the underlying construct due to this broad construct origin criteria. They suggest that a conceptual model that explains the underlying structure of a latent variable and drives the instrument development is necessary. Further they argue that a theory-driven approach to determining the relationship between the conceptual model and the score on the measure is essential. That the COSMIN guidelines do not consider whether measures are developed based on a conceptual model is a concern when measuring meaning in life. As I have noted above, concepts within eudemonic wellbeing lack definitional specificity and it is particularly important that a systematic review of meaning in life establishes whether a precise conceptual framework has been used to develop

a scale. Further, the lack of an appropriate evaluation of the relationship between the conceptual model and the structure of the instrument suggests an insufficiency in the COSMIN guidelines.

However, if we set aside the overly loose approach that COSMIN takes in evaluating the underlying constructs, we can consider the evidence provided on structural validity (whether instrument scores accurately reflect the factor structure of an underlying conceptual model [Mokkink et al., 2011]) in these studies. Examining the dimensionality of instruments may provide further understanding of the underlying structure of meaning in life. If we do consider structural validity of these instruments within the framework of the COSMIN guidelines, four measures showed sufficient evidence for structural validity as determined by either confirmatory or exploratory factor analyses. The aim of an exploratory factor analysis (EFA) is to "identify the common factors that explain the order and structure among measured variables" (Watkins, 2018, p. 220) whereas a confirmatory factor analysis tests this relationship between measurable variables and either latent variables or factors as determined by a hypothesis a priori (Jackson et al., 2009). The evidence regarding the underlying structure of the measures was inconsistent. Three of the measures exhibit evidence for a one factor structure (Salsman et al., 2020; Wang et al., 2015; Shulenberg et al., 2011) which means that these instruments consist of one underlying factor. However, one instrument exhibited a five-factor structure (Xia et al., 2017) which suggests this instrument has a more complex structure. Two of these used a CFA (Salsman et al., 2020; Schulenberg et al., 2011) while two used EFA (Wang et al, 2015; Xia et al., 2017). These inconsistent results may be due to many factors.

On one hand, it may be that the different underlying structures are representative of a true difference in the experience of meaning in life in different cultural settings or within different populations. This difference may be a true reflection of variability in the underlying structure of the construct, which is what has caused the different factor structures of the instruments. On the other hand, this inconsistency may be due to differences in the

development of the instruments. When examining the factor structure of an instrument, the construct is not measured directly. Instead, instruments are indirect measures of the latent variables. Many researchers aim to create a unidimensional instrument, as these are thought to exhibit conceptual homogeneity (Furr, 2013). Conceptual homogeneity means that all items represent the same underlying construct. An instrument which is not unidimensional risks including overlapping constructs and when the aim is to measure one construct, this can cause contamination from other similar constructs. This means that a score on an instrument is measure of multiple constructs and is at risk of becoming meaningless as it is unclear what it measures. This is a likely risk in the context of eudemonic wellbeing, which suffers from enmeshed constructs. A unidimensional instrument avoids this risk. However, certain instruments may also include an overly narrow focus on meaning in life, which may cause criterion deficiency. This means that some aspects of meaning in life may be ignored. Certainly, these inconsistent factor analyses need to be examined further to understanding whether they are a function of the structure of instruments of meaning in life or meaning in life itself.

Further, some measures expanded the concept of meaning in life. For example, an instrument was developed to include need for meaning, meaning confusion, meaning avoidance, and meaning anxiety (Zhang et al., 2018) and another included reflectivity as a part of meaning in life (Hill et al., 2019). Including new ways of understanding meaning in life may help develop a comprehensive definition, but also introduces further variability in our understanding of the construct. Also, there was an attempt to develop a way of measuring meaning in life for specific populations, such as those impacted by HIV/AIDS (Audet et al., 2015). This suggests that perhaps meaning in life is not a singular construct which is experienced and presents identically in different populations. Instead, it may be an adaptable construct, which can refer to different aspects in different contexts and populations, as suggested by Leontiev (2013). Different populations and cultures may exhibit different patterns

of meaning in life. If meaning in life changes depending on where it is measured, instruments may not be applicable in varied contexts.

There was a concern about the lack of generalizability of the studies. Five of the measures reviewed in this article were developed in North America, four in East Asia and one in Europe. The previous systematic review on this subject, reported measures developed in Europe, North America, China, Australia, and Israel (Brandstätter et al., 2012). There continue to be a lack of measures developed in South America, Africa, and South Asia. Measures of meaning in life have subsequently been validated among culturally diverse populations such as among a Hausa-speaking Internally Displaced Population in Nigeria (Chika Chukwuorji, 2019), in Chile (Steger & Samman, 2012) and among Hindi speaking populations in India (Singh et al., 2016). While this suggests the validated measures may be appropriate for use in populations they were not developed in, aspects of meaning in life that are unique to these cultures may be lacking. This may mean the measures only capture a part of the experience of meaning in life. Further, only one study assessed cross-cultural validity (Salsman et al., 2020). This was assessed within different groups (sex, race, education, age) of a population who were all living in the USA and able to complete the questionnaire in English. Therefore, it did not provide evidence for validity among different regions or cultures of the world.

All but one of the studies was developed based on classical test theory whereas one used classical test theory in combination with item response theory (IRT) in its development and validation (Salsman et al., 2020). IRT can be used to draw more sophisticated conclusions regarding the properties of individual items whereas CTT is generally used to examine latent variable models. McKenna and Heaney (2021) suggest that CTT is limited in the conclusions that can be drawn from it, in ways that IRT and Rasch Measurement Theory (RMT) are not. This suggests that relying on instruments which have only been evaluated based on CTT is limited in the information that can be provided about the instruments, as no item specific

information is available. However, others have argued that CTT is not worse than IRT, only that different conclusions can be drawn using these two methods (Raykov & Marcoulides 2016). In the evaluation of these measures there is information that is missing that could be evaluated using further sophisticated data analyses.

A previous systematic review focusing on assessment instruments of measures of life was published in 2012. Similarly, to the current review, the authors did not suggest a particular measure to be better in all contexts, but instead encouraged researchers to choose a measure that is appropriate for the aims of their specific study. What the authors also noted was the challenge in defining meaning in life, and that many studies offered vague definitions, which was in accordance with the current review. In comparison to the current review, the previous systematic review did not determine whether the measures were of good enough quality to be recommended for use, as they did not use a scoring system such as Cosmin. This leads to some difficulties in comparing the overall quality of the measures in the two reviews. Generally, researchers need to choose a measure based on their specific requirements and the aim of their study. Recommending a specific measure to be used in all research on meaning in life is not appropriate.

The aim of this systematic review was to critically evaluate instruments that measure meaning in life As has been expanded upon in the discussion section, instruments were poorly developed and exhibited a lack of evidence for valid and reliable conclusions that can be drawn. As a consequence of this, along with inconsistent evidence regarding the structure of the instruments, this systematic review of meaning in life has revealed shortcomings in the existence of high-quality assessment instruments of meaning in life.

Limitations

There are some significant limitations to the present review which are important to note. Firstly, one reviewer completed all parts of the systematic review with a second reviewer only

rating a portion of the studies. Secondly, only two databases were searched for applicable studies. Thirdly, despite the COSMIN guidelines being recommended for use in this context, the COSMIN guidelines have received criticism as their ability to support high quality research has been questioned and they may be inappropriate for use in this context. I will review these factors and their impact on the present review.

Recommendations by the JBI Manual for Evidence Synthesis (Aromataris & Munn, 2020) and the COSMIN guidelines (Prinsen et al., 2018) state that at least two reviewers should collaborate on the systematic review. When multiple reviewers collaborate, it is possible to measure the agreement between them. This gives an indication of the reliability of the results. This systematic review was completed by the author with a second reviewer rating only a portion of the results. There was high agreement between the two reviewers, which suggests that the results are reliable. However, the lack of a second reviewer to assess all steps taken in this review is a significant limitation that needs to be considered when reviewing the results.

A further limitation of the current systematic review is that only two databases were searched. The results from these two databases (psycinfo and MEDLINE) and the key terms used provided around 2,000 articles. Due to time constraints, no further databases were searched. The objective of the systematic review was to provide a comprehensive overview of the state of the literature, and it is possible that there are articles that fit the inclusion criteria that were not present in the search, due to not being present in these two databases. However, no further developed scales were found by trailing the reference lists of the articles. Therefore, a possible limitation of the present review is that it does not provide a comprehensive review of all newly developed scales.

Further, the use of the COSMIN guidelines themselves are a limitation of the current study. The COSMIN guidelines were developed to assess patient-reported outcome measures and aid the "selection of health outcome measurement instruments in research and clinical

practice." A patient-reported outcome is "any report of the status of a patient's health condition that comes directly from the patient without interpretation of the patient's response by a clinician or anyone else" (FDA, 2009). On one hand, meaning in life can be an outcome in clinical research regarding both physical health (such as cancer) and mental health conditions (depression). On the other hand, meaning in life is also a broad transdiagnostic construct relevant to non-clinical populations. The present and a previous systematic review found measures of meaning in life were developed both in undergraduate 'healthy' samples as well as clinical samples (Brandstätter et al., 2012). Meaning in life as a construct holds a dual role as it is related to mental wellbeing in the general population but also relevant to clinical research and practice. Therefore, meaning in life can take the form of a patient-reported outcome but in other contexts it may not be appropriate to refer to it as a patient-reported outcome, because it is being measured in non-health related outcomes in a non-clinical population. Therefore, the COSMIN guidelines, which were specifically developed for health outcome instruments, may be appropriate for certain instruments in this review, but not for the use of other instruments, depending on populations whom they consider. To further complicate this, measures of meaning in life have been developed to assess multiple factors which could contribute towards it, such as evaluating the presence and sources of meaning in life, but also crisis and search for meaning in life (Brandstätter et al., 2012). The breadth of the applicability of the concept 'meaning in life' means guidelines which have been developed to evaluate measures with specific aims or populations are not relevant for other measures, which focus on alternative aims or populations. The COSMIN guidelines may be more appropriate for certain measures of meaning in life, such as those developed in clinical populations to evaluate interventions, than others, which focus more on a non-clinical population. A further evaluation would be necessary to determine which instruments were developed so they were appropriate to be evaluated as patient-reported outcome measures.

A further criticism of the COSMIN guidelines is how they evaluate content validity. To evaluate content validity, the COSMIN guidelines assesses whether the items that make up an instrument are relevant, comprehensive, and comprehensible (Prinsen et al., 2018). McKenna and Heaney (2021) argue that this way of evaluating content validity does not appropriately appraise the underlying construct as it only requires that patients, professionals and the reviewer consider whether the contents of the measure are relevant, comprehensive, and comprehensible. They argue that this is not enough to determine the validity of a measure. It is difficult to determine whether the items meet these criteria due to the confusion around meaning in life, which could diminish the value of the evaluation of content validity in this review. A further study should consider the complexity in the relationship between the underlying construct and the measure as an important part of evaluating measures.

An additional limitation of the COSMIN guidelines that McKenna and Heaney (2021) note, is that COSMIN is solely based on opinion rather than developed based on an evidence base. While it is based on the collective opinions of experts within the field, they are experts within health-related quality of life. The experience within health-related quality of life may mean that COSMIN guidelines are poorer quality or perhaps even unsuitable for evaluating other outcomes. These guidelines may not be of adequate quality or appropriate or relevant for the current review. Consequently, McKenna and Heaney (2021) have noted that the COSMIN guidelines lack validity and reliability, as different reviews of the same measures have reached different conclusions. This may be due to the fact reviewers with different experience and understanding of measurement theory use the COSMIN guidelines differently (McKenna and Heaney, 2021). In the current study this is an especially relevant limitation, as only one reviewer was involved.

Multidimensional instruments are a further weakness in the validation of these measures. There is a lack of appropriate ways of evaluating composite measures – measures

which are made up of more than two separate unidimensional scales (McKenna & Heaney, 2021). It is unclear how to judge how separate scales are related to each other. It is not appropriate to add together separate scales to a composite score as it is unclear what this score means (McKenna & Heaney, 2021). In this systematic review they were assessed separately, however, in practice the use of them will be combined,

Conclusion

Ten newly developed measures which were published since the previous systematic review of meaning in life measures were assessed using the COSMIN guidelines. The development of all included measures was found to be of inadequate quality. A significant part of the inadequate quality assessment was due to a lack of piloting or cognitive interviewing in the development of the measures. However, the COSMIN guidelines used have been criticised and may not be appropriate in the evaluation of these measures. The aim of this systematic review was to determine whether any measures of good quality could be recommended and provide a way to conceptualise meaning in life. However, due to the poor quality of the measures, no measures could be recommended or used for this purpose. The development of high-quality measures should be a priority to ensure meaning in life can be evaluated appropriately.

Integration

Overview

This doctoral thesis combined an evaluation of instruments that measure meaning in life, published in the last decade, with a dimensionality analysis of an instrument (MMM) which measures meaning in life. This has highlighted the necessity of good quality methodology in the development of scales along with the importance of critically evaluating the relationship between a construct and its assessment. This is especially important considering the context of a reproducibility crisis currently affecting psychology (Lilienfeld, 2017) and a move within mental health services within the NHS away from a biomedical model towards a recovery-based model.

Context of work

Currently, psychology is faced with a reproducibility crisis which is due to the fact that psychological science is less reproduceable than expected (Lilienfeld, 2017). For example, one study which aimed to replicate 100 original studies found that only a minority of the replications were statistically significant (Open Science Collaboration, 2015). One factor that may contribute to the reproducibility crisis is how researchers measure psychological constructs. A popular way of measuring psychological constructs is by using self-reported scale scored based on a Likert scale, which was a focus of both my empirical study and systematic review. However, self-report scales have been criticised and may be a reason for the lack of consistent results, and therefore contributed to the reproducibility crisis.

During this work, I have come to appreciate the time and effort that is required to develop, validate, and evaluate measures of adequate quality. In light of this, it is perhaps understandable that the systematic review highlighted poor quality of developed scales. It is unknown how many of these scales have been used in further research, considering at least 59 instruments were developed before 2012 (Brandstätter et al., 2012). While this article has

highlighted concerns regarding the quality of the development of measure of meaning in life it is likely that these concerns regarding meaning in life scales are not unique to this topic but is a pervasive challenge in psychological research. It has been argued for a long time that reliability and validity of self-report questionnaire is affected by how questions are written and presented to participants, and that cognitive testing is a necessary part of the development of scales despite the minority of newly developed scales having undergone such testing (Krosnick, 1999). In the development of measures, we need to be aware and follow methodological recommendations.

Other researchers have suggested that improving the quality of the development of such self-report scales is not sufficient (Uher, 2022). Instead, the choice to use rating scales is based on ease and efficiency of producing research rather than producing meaningful results (Uher, 2022). Uher argues rating scales are developed based on overarching fallacies such as conflating what we aim to measure with the way it is measured (Uher, 2021). We need to examine the relationship between response patterns and the phenomena in question itself as currently we are unable to provide such evidence. She further argues that the response to the reproducibility crisis within psychology has been the evaluation and development of better methods of data analysis, rather than data generation (Uher, 2021). As we develop more sophisticated statistical methods, such as Exploratory Graph Analysis (EGA), we can assess response patterns and understand how items are related to both each other and the scale. However, there is a danger in relying on data analysis without also including participants and those who will be using the questionnaire in its evaluation and development. Uher (2021) further argues that increased sophistication of data analysis does not negate the need to ensure that data generation is appropriate.

Other researchers have argued that people can use numerical scales in a meaningful way (Kaiser & Oswald, 2022). They showed that integers people chose to describe feelings

can predict behaviours, and that this relationship is close to linear. Also, the number has more predictive power than combined socio-economic data. This consistency between numerical scales and actions suggests the reliability of translating feelings into numerical values (Kaiser & Oswald, 2022). While the development of a scale is more complex as it includes writing and choosing a combination of questions, this research suggests that scales themselves can be meaningful and that the development of a good-quality measures is attainable. This thesis highlights concerns about the poor quality of instrument development and the value that novel statistical techniques in conjunction with good quality methodology can accomplish.

Extent of synergy achieved

There is a clear interaction between the systematic review and the empirical article, and they complement each other. While the systematic review took a broad view of the measurement of meaning in life by examining the quality of several scales developed to measure meaning in life, the empirical study added a piece of knowledge to the evidence base regarding one specific measure of meaning in life. There are similarities between these two projects. Both consider the measurement of meaning in life and require us to acknowledge the complexity of this construct. Both also consider scale construction and evaluation, albeit from slightly different perspectives. A significant difference between the two studies is that the empirical study added novel information to the evidence base regarding measurement of meaning in life. This new information helped us refine an existing scale of meaning in life and helps us improve our understanding of people's experiences of meaning in life. In conclusion, the empirical study and the systematic review contribute to the evidence base regarding meaning in life as a construct. While the studies differ in their methodology and specific aims, they both provide further information regarding the measurement of meaning in life.

While I worked on both my systematic review and my empirical study simultaneously, my systematic review provided a conceptual basis for the empirical study. Firstly, undertaking my systematic review helped me develop a better understanding of meaning in life and its measurement. Evaluating these studies helped me understand the varied definitions, and the methodology used. Also, I gained an overview of the similarities and differences in how different researchers understood meaning in life and their aims in measuring meaning in life. This helped me clarify my understanding of meaning in life and its assessment which was helpful in making sense of the empirical study.

Secondly, the process of completing the systematic review allowed me to develop an overview and understanding of scale development and quality evaluation. Especially, since I used the COSMIN guidelines, which offer a standardised way of assessing responsiveness, construct and criterion validity, reliability, cross-cultural validity, internal consistency, structural validity, and content validity. This helped me better understand these aspects of scale construction and evaluation. Using the COSMIN guidelines also helped me understand the importance of considering the specific aims of my study. This was especially helpful as I had not been part of the development of the MMM, only its evaluation. A further way the systematic review provided a foundation for my empirical study was the need for me to develop my knowledge about how to evaluate different statistical outcomes. This includes more practical knowledge such as how confirmatory factor analyses are evaluated, and what the cut-offs are for determining validity.

Also, through the systematic review I gained an understanding of the hypothesis and results of dimensionality of newly developed scales. I was able to assess how the structure of the different scales compared. I was also able to compare any differences or similarities to the dimensionality of the MMM. This helped me gain further understanding of how others have assessed the structure of meaning in life.

Challenges and dilemmas during the project

Systematic review

One of the main challenges with the systematic review was determining its scope. Initially I was planning on including both articles which describe the development of articles along with any validation studies regarding meaning in life scales published during the same timeframe. Initially, I gathered both types of articles but once I started evaluating the development studies, I realised it would not be possible to evaluate the validation studies to a standard that would be appropriate for a doctoral level project I had to make the decision to not include them. At this time, I had already spent considerable time on the validation studies. This taught me that it is more important to ensure the quality of your work, rather than embarking on a project that is not feasible. If I had attempted to scope out the work necessary for each type of article before starting, I could have saved myself unnecessary labour.

Another challenge in this systematic review was that, on one hand, I wanted to update a previous systematic review. This meant that I wanted to use similar wording and use similar methodology to ensure continuity. On the other hand, considering that the previous systematic review was published over ten years ago, I also wanted to include the most recent evidence base that exists to ensure this systematic review was up to date. I resolved this dilemma by ensuring that I included search terms from the original article, however, I assessed the included articles based on the most research evidence base. This way I was able to combine factors based on each priority. This made me realise how every choice I made would have repercussions and that there was not a perfect way of managing the situation. Instead, I attempted to make a conscious choice that I could defend at each stage while being mindful that other researchers – especially those with more time and resources – may have made a different choice.

Empirical study

The most significant challenge of my empirical study was my lack of experience of complex statistical processes. Much of my time (especially in the early stages of my work on this thesis) was spent learning to understand EGA and network analysis. This was a combination of understanding the theoretical underpinnings of both latent variable models and the newer network models. This was further complicated by the fact that EGA can only be done within R studio. This required me to learn how to code, which I had never attempted before. All my previous experience in statistical analysis was within SPSS, which is different from R. While I only learnt to use R for EGA and the associated tasks, this still required me to spend a lot of time making sense of it. With the support of my supervisor, I was able to learn this skill. This taught me the importance of allowing time to make sense of complex matters and the necessity of not trying to rush one's understanding. I also appreciated using such a novel process, as the specific type of EGA which was used in this study had not been published when I started this project in 2021. Realising this study will therefore be a novel contribution to this field makes the project feel more meaningful.

The inclusion of service users within this project was a challenge that I did not resolve. We should strive to include service users in the research efforts. On one hand, service user involvement is a valuable activity which can shape the direction of research (Tait & Lester, 2005). On the other hand, a discrete study, such as this one, did not have space for changes in direction. It was also difficult to think of a way of involving service users in a meaningful, rather than tokenistic way. An understanding of complex statistical formulations would be necessary to ensure co-production. The lack of service user involvement within this study reflects the lack of service user involvement in the development of measures which this thesis has highlighted. A further reason service user involvement was challenging was that I found it difficult to determine a specific audience. Meaning in life is a transdiagnostic concept which is relevant to all people. A service user may not find this relevant to them,

while someone who had not used mental health services may find this topic personally relevant. It may have been valuable to recruit individuals with personal interest in meaning in life and experiences such as mental ill-health, trauma, or other difficult experiences to consider the implications of the results.

Additional reflections

Reflecting on this work makes me consider its impact on me as a future psychologist. This empirical study was my first independent piece of research at this high level. After I qualify, I will work in an NHS service in a fully clinical role. I am tempted to consider my research career having concluded with the publication of this thesis as I will not have an opportunity to work as a researcher in the future. However, I have considered the importance of continuing to identify as a scientist-practitioner after qualification. This following definition of the essence of being a scientist practitioner helped me consider how this would be possible:

> The scientist-practitioner model is… a harmonious balance of both research and practice. This balance does not necessarily indicate that individuals will spend equal time in each arena or that they will be equally proficient in both domains. Instead, it provides for a dynamic understanding and influence between a psychologist's research and practice. (Horn et al., 2007, p.809)

This way of defining a scientist-practitioner helped me realise that despite my upcoming move to a clinical role, I can still embody the scientist-practitioner model in the future. This thesis has helped me become a critical consumer and researcher. In terms of evaluating and applying research, my work on this project has helped me be more critical of research. Prior to conducting independent research, I would rely on research being peer-reviewed to trust its claims. Now that I have conducted and evaluated research, I have gained confidence and skills in critiquing research methodologies. This is especially true in my future choices

around what techniques and methods I will use for measurement. As a qualified psychologist, despite working in a clinical setting, I will engage in research projects on smaller scales such as audits, along with using a hypothesis-based approach with clients. In summary, this project has helped me appreciate the importance of being a scientist-practitioner in my career as a psychologist within the NHS.

Impact

Key Messages

1. This systematic review of measures of meaning in life found that the development of the ten scales identified exhibited inadequate quality. This implies that the development of these scales of meaning in life need to be further developed and refined. Another possibility is that novel measures which capture the complexity of meaning in life are developed. Both clinicians and researchers should be mindful of the poor quality of the development of these scales when using them in their clinical practice or within research studies. Being mindful of these limitations ensures that they are considered in decision making or when assessing the impact of clinical practice or research studies.
2. The empirical study used a diverse sample to evaluate the dimensionality of the MMM (a measure of meaning in life). This measure was characterised by four separate dimensions which were loss of meaning in life, purpose, integration of circumstances, and high-level action identification. Determining the dimensionality of this measure suggests that meaning in life is a complex construct which includes multiple cognitive and affective processes. Delineation of these dimensions offer better understanding of meaning in life (as conceptualised in this scale) and its measurement for clinicians and researchers. This scale offers a nuanced approach to

assessing meaning in life which can be applied to both clinical and research-based aims. The ultimate choice will depend on the specific research questions and clinical targets.

Academic Impact

Scale construction

The empirical study used a novel technique to manage wording effects in a scale made up of both negatively and positively worded items. Arguments have been made for and against including items of different valence in the same scale. Certain researchers argue that there are advantages to including a mix of wording in scales as including only positively worded items increases the possibility of acquiescence bias in people's responses (Solís Salazar, 2015). Other researchers argue that mixed wording of scales may reduce the meaningfulness of responses as participants fail to consider the valence of the item, thereby not understanding the question asked (Steinmann et al., 2022). This study adds to the evidence base regarding the potential to include both negatively and positively worded items in scale construction by considering the wording effect.

Psychometrics

A unique aspect of the empirical study is the use of random intercept EGA (riEGA), which is an adaptation of EGA that takes into account wording effects. To my knowledge, this study is one of the first ones to use this technique to determine the dimensionality of a scale. Using such new techniques functions as a way of assessing their effectiveness. This allows us to evaluate the effectiveness of such a method in a real-world data, compared to a simulated data set. Use of novel techniques also encourages researchers to refine and develop these techniques. As these techniques are further refined their use will become more popular and their effectiveness will increase.

A further contribution of this study was the comparison between latent variable and network analysis models. While they were developed based on different assumptions and data generation models, they have been found to be mathematically equivalent. This study further suggests that structures develop based on a network analysis model can be compared and thus show good fit. This provides researchers with further evidence to use a combination of methods that fit best with their research aims.

Clinical Impact

Mental health services in the UK are moving away from the biomedical model towards a recovery-based model (Department of Health, 2011). Evaluating measures of meaning in life has provided caution of the use of measures of meaning in life due to poor methodology used in their development. The evaluation of the dimensionality of the MMM (measure of mundane meaning) provides a way of assessing meaning in life from the perspective of this measure and its conceptualisation of meaning in life. Being able to measure the different aspects of meaning in life and understand how they relate to each other may help produce more person-centred interventions within clinical psychology. This can help services move from symptom-based outcome measures towards meaning and recovery enhancing ways of measuring the impact of clinical work.

Societal Impact

The Organisation for Economic Co-operation and Development (OECD, 2013) has suggested that all its member countries (which includes the UK as part of the G7 among its 38 member states) measure subjective wellbeing on a national basis. This presents a shift away from using gross domestic (GDP) as a singular way of measuring development and progress. Including subjective wellbeing paints a more complex and thorough picture of the progress of a nation compared to GDP which only assesses the economy. However,

subjective wellbeing is less objective than GDP, and due to the recency of including such measures their development is still in progress.

The OECD suggests three aspects of subjective wellbeing comprising of life evaluation (evaluative well-being), affect (experiential well-being) and eudaimonia, which they define as a sense of meaning and purpose in life, or good psychological functioning. There are agreed-upon ways of measuring both evaluative and experiential, but eudaimonia does not have a clear conceptualisation within this context (Martela & Ryan, 2023). Instead, eudaimonia has two competing operationalizations that of either psychological functioning or meaning in life (OECD, 2013). Some researchers argue that psychological functioning should be favoured over meaning in life, and that meaning in life should instead be considered an aspect of evaluative wellbeing (Martela & Ryan, 2023). However, with the understanding of meaning in life developed through this thesis, I will argue that meaning in life provides a framework that best describes what we understand as eudaimonia. While it is clear there are many poorly developed scales of meaning in life, this thesis provides a way of conceptualising meaning in life. This conceptualisation also described how meaning in life is related to both more cognitive (through integration of circumstances) and behavioural (through high-level action identification) aspects. In conclusion, the impact of this study is the provision of a way of conceptualising and measuring meaning in life to provide a clear way forward in measuring subjective wellbeing on a national scale.

Dissemination

I have two approaches for the dissemination of my research. Firstly, I want to attempt to disseminate my work outside of an academic context. Within the course a presentation of the empirical study will take place, which will be attended by other doctorate students and staff. This will require the adaptation of the empirical study to a presentation, which can also be disseminated. This presentation could be in a clinical context, such as in services for

people with severe mental illness. Another possibility could be palliative or psycho-oncology services where meaning in life is often discussed with patients. Another way of disseminating this material is through social media and blogs. For example, the "National Elf Service" is a popular blog which produces short summaries of academic work for anyone who is interested in mental health in the UK.

I also plan on submitting my empirical study to a journal. Finding an appropriate article to submit to is challenging as deciding on a journal for this study is difficult. There are three main topics that I could choose the journal based on. Firstly, a journal which is focused on meaning in life. This may be a journal such as the Journal of Happiness studies which has published many articles about meaning in life. Specifically, two articles included in my systematic review were published in this study. On the other hand, I could also focus on journals which focus on psychometrics considering the novelty of riEGA. I believe this study is one of the first to use riEGA, which might make the topic more interesting for researchers who may be considering using riEGA in developing their questionnaires. Thirdly, the initial article presenting the development of the MMM was published in the International Journal of Cognitive Therapy. This was due to meaning in life being presented as a transdiagnostic construct relevant in cognitive therapy. Considering these options, I also want to reflect on my personal values. If possible, I would like to support journals which provide open access and do not require a costly subscription. Personally, I want to enable people to be able to access high quality and peer reviewed research, which is not dependent on payment. In summary, there are many journals that are suitable for this empirical study.

Similarly, my systematic review does not necessarily neatly fit one journal. The previous systematic review on measures of meaning in life was published in the journal Psycho-Oncology. Meaning in life is a relevant construct within palliative care. Palliative care is a research area within cancer so one can see why this article was published in this

journal. The articles included in the systematic review were published in diverse journals such as Quality of Life Research, Journal of Clinical Psychology, and the Journal of Positive Psychology. To be of publishable standard, it would require a second reviewer to assess all articles, as currently only a few of them are reviewed by two people.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition* (DSM-5). Washington, D.C.: American Psychiatric Association.

Alexandrova, A. & Fabian, M. (2022). *The science of wellbeing*. John Templeton Foundation.

Anthony, W. A. (1993) Recovery from mental illness: The guiding vision of the mental health system in the 1990s. *Psychosocial Rehabilitation Journal,* 16, 11-23.

Antonovsky, A. (1979). *Health, stress and coping.* San Francisco: Jossey–Bass.

Aromataris, E., & Munn, Z. (2020). *JBI Manual for Evidence Synthesis.* Available from https://synthesismanual.jbi.global. https://doi.org/10.46658/JBIMES-20-01

Audet, C. M., Wagner, L. J., & Wallston, K. A. (2015). Finding meaning in life while living with HIV: Validation of a novel HIV meaningfulness scale among HIV-infected participants living in Tennessee. *BMC Psychology*, *3*(1), 15. https://doi.org/10.1186/s40359-015-0070-7

Baumeister, R. F. (1991). *Meanings of life.* New York: Guilford Press.

Bjelland, I., Dahl, A. A., Haug, T. T., & Neckelmann, D. (2002). The validity of the Hospital Anxiety and Depression Scale: An updated literature review. *Journal of Psychosomatic Research, 52*(2), 69–77. https://doi.org/10.1016/S0022-3999(01)00296-3

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 10*, P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology*, *9*, 91–121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Brandel, M., Vescovelli, F., & Ruini, C. (2017). Beyond Ryff's scale: Comprehensive measures of eudaimonic well-being in clinical populations: A systematic review. *Clinical Psychology & Psychotherapy, 24*(6), 1524–1546. https://doi.org/10.1002/cpp.2104

Brown, G. P, Roach, A., Irving, L., & Joseph, K. (2008). Personal meaning: A neglected transdiagnostic construct. *International Journal of Cognitive Therapy,1*, 223-236. https://doi.org/10.1521/ijct.2008.1.3.223.

Brandstätter, M., Baumann, U., Borasio, G. D., & Fegg, M. J. (2012). Systematic review of meaning in life assessment instruments. *Psycho-Oncology*, *21*(10), 1034–1052. https://doi.org/10.1002/pon.2113

Chika Chukwuorji, J., Ekpedoho, E.A., Ifeagwazi, C.M., Iorfa, S.K., & Nwonyi, S.K. (2018). Psychometric properties of the Meaning in Life Questionnaire – Hausa version among internally displaced persons in Nigeria. *Transcultural Psychiatry, 56,* 103-122. https://doi.org/10.1177/1363461518794218

Christensen, A. P., & Golino, H. (2019). Estimating the stability of psychological dimensions via bootstrap exploratory graph analysis: A Monte Carlo simulation and tutorial. *Psych, 3*(3), 479-500. https://doi.org/10.3390/psych3030032

Christensen, A. P., Garrido, L. E., & Golino, H. (2020a). Comparing community detection algorithms in psychological data: A Monte Carlo simulation. *PsyArXiv.* https://doi.org/10.31234/osf.io/hz89e

Christensen, A. P., Garrido, L. E., & Golino, H. (2020b). Unique variable analysis: A novel approach for detecting redundant variables in multivariate data. *PsyArXiv.* https://doi.org/10.31234/osf.io/4kra2

Christensen, A. P., Golino, H., & Silvia, P. J. (2020c). A psychometric network perspective on the validity and validation of personality trait questionnaires. *European Journal of Personality, 34,* 1095–1108. https://doi.org/10.1002/per.2265

Christensen, A. P., & Golino, H. (2021). On the equivalency of factor and network loadings. *Behavior research methods,* 53(4), 1563–1580. https://doi.org/10.3758/s13428-020-01500-6

Cramer, A. O. J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S.H., Kendler, K.S. & Borsboom, D. (2012). Dimensions of normal personality as networks in search of equilibrium: You can't like parties if you don't like people. *European Journal of Personality, 26,* 414-431. https://doi.org.ezproxy01.rhul.ac.uk /10.1002/per.1866

Crumbaugh J. C. (1968). Cross-validation of Purpose-in-Life test based on Frankl's concepts. *Journal of Individual Psychology, 24*(1), 74–81.

Deacon, B. J. (2013). The biomedical model of mental disorder: A critical analysis of its validity, utility, and effects on psychotherapy research. *Clinical Psychology Review, 33*(7), 846–861. https://doi.org/10.1016/j.cpr.2012.09.007

Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: An introduction. *Journal of Happiness Studies, 9*(1), 1–11. https://doi.org/10.1007/s10902-006-9018-1

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of Research in Personality, 19,* 109–134.

Department of Health. (2011). *No Health Without Mental Health: A cross-government mental health outcomes strategy for people of all ages.* https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213761/dh_124058.pdf

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97*(2), 143–156. https://doi.org/10.1007/s11205-009-9493-y

Division of Clinical Psychology. (2001). *The Core Purpose and Philosophy of the Profession*. The British Psychological Society.

Fabrigar, Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi.org/10.1037/1082-989X.4.3.272

Farkas, M., Gagne, C., Anthony, W., & Chamberlin, J. (2005). Implementing recovery oriented evidence based programs: Identifying the critical dimensions. *Community mental health journal, 41*(2), 141–158. https://doi.org/10.1007/s10597-005-2649-6

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin, 99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Frankl. (2004). *Man's search for meaning : the classic tribute to hope from the Holocaust*. Rider.

Freud, S. (1930). *Civilization and its Discontents*. W. W. Norton & Company.

Furr, R. M. (2013). *Scale construction and psychometrics for social and personality psychology*. (2nd edition). SAGE Publications Ltd. https://doi.org/10.4135/9781446287866.n1

Garcia-Padina, A., Abad, F., & Christensen, A., Golino, H., & Garrido, L. (2022). Dimensionality assessment in the presence of wording effects: A network psychometric and factorial approach. https://doi.org/10.31234/osf.io/7yqau.

George, L. S., & Park, C. L. (2016). Meaning in life as comprehension, purpose, and mattering: Toward integration and new research questions. *Review of General Psychology, 20*(3), 205–220. https://doi.org/10.1037/gpr0000077

George, L. S., & Park, C. L. (2017). The Multidimensional Existential Meaning Scale: A tripartite approach to measuring meaning in life. *The Journal of Positive Psychology, 12*(6), 613–627. https://doi.org/10.1080/17439760.2016.1209546

Golino, H. F., & Epskamp, S. (2017). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS ONE,* 12(6), e0174035. https://doi.org/10. 1371/journal.pone.0174035

Golino, H., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Sadana, R., Thiyagarajan, J. A., & Martinez-Molina, A. (2020). Investigating the performance of exploratory graph analysis and traditional techniques to identify the number of latent factors: A simulation and tutorial. *Psychological Methods, 25*(3), 292–320. https://doi.org/10.1037/met0000255

Golino, H., Christensen, A., Moulder, R., & Garrido, L. E. (2021a). *EGAnet: Exploratory Graph Analysis—a framework for estimating the number of dimensions in multivariate data using network psychometrics* (1.0.0) [Computer software]. Retrieved from https://CRAN.R-project.org/package=EGAnet.

Guerra, M. P., Lencastre, L., Silva, E., & Teixeira, P. M. (2017). Meaning in life in medical settings: A new measure correlating with psychological variables in disease. *Cogent Psychology, 4*(1), 1286747. https://doi.org/10.1080/23311908.2017.1286747

Heintzelman, S. J., & King, L. A. (2014). Life is pretty meaningful. *American Psychologist, 69*(6), 561–574. https://doi.org/10.1037/a0035049

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and*

Psychological Measurement, *66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Hill, P. L., Burrow, A. L., Sumner, R., & Young, R. K. (2015). Life is pretty meaningful and/or purposeful?: On conflations, contexts, and consequences. *American Psychologist*, *70*(6), 574–575. https://doi.org/10.1037/a0039063

Hill, C. E., Kline, K. V., Miller, M., Marks, E., Pinto-Coelho, K., & Zetzer, H. (2019). Development of the meaning in life measure. *Counselling Psychology Quarterly*, *32*(2), 205–226. https://doi.org/10.1080/09515070.2018.1434483

Horn, R. A., Troyer, J. A., Hall, E. J., Mellott, R. N., Coté, L. S., & Marquis, J. D. (2007). The Scientist-Practitioner model: A rose by any other name is still a rose. *American Behavioral Scientist*, *50*(6), 808–819. https://doi.org.ezproxy01.rhul.ac.uk/10.1177/0002764206296459

Huta, V., & Waterman, A. S. (2014). Eudaimonia and its distinction from hedonia: Developing a classification and terminology for understanding conceptual and operational definitions. *Journal of Happiness Studies*, *15*(6), 1425–1456. https://doi.org/10.1007/s10902-013-9485-0

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting Practices in Confirmatory Factor Analysis. *Psychological Methods*, *14*(1), 6–23. https://doi.org/10.1037/a0014694

Janoff–Bulman, R. (1992). *Shattered assumptions: Towards a new psychology of trauma.* New York: Free Press.

Jung, C. G., Jaffe, A., Winston, C., & Winston, R. (1989). *Memories, Dreams, Reflections* (Revised ed.). Vintage Books.

Kam, C. C. S. (2018). Why do we still have an impoverished understanding of the item wording effect? An empirical examination. *Sociological Methods & Research, 47*(3), 574–597. https://doi-org.ezproxy01.rhul.ac.uk/10.1177/0049124115626177

Kaiser, C., & Oswald, A. J. (2022). The scientific value of numerical measures of human feelings. *Proceedings of the National Academy of Sciences of the United States of America, 119*(42), e2210412119. https://doi.org/10.1073/pnas.2210412119

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika 39,* 31–36.

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology*, 67, 332–339.

Kazdin, A. E. (2008). Evidence-based treatment and practice: New opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. *American Psychologist, 63,* 146–159. https://doi.org/10.1037/0003-066X.63.3.146

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior, 43*(2), 207–222. https://doi.org/10.2307/3090197

King, L. A., & Hicks, J. A. (2021). The science of meaning in life. *Annual Review of Psychology, 72*(1), 561–584. https://doi.org/10.1146/annurev-psych-072420-122921

King, L. A., Hicks, J. A., Krull, J. L., & Del Gaiso, A. K. (2006). Positive affect and the experience of meaning in life. *Journal of Personality and Social Psychology, 90*(1), 179–196. https://doi.org/10.1037/0022-3514.90.1.179

Kjaergaard, M., Arfwedson Wang, C. E., Waterloo, K., & Jorde, R. (2014). A study of the psychometric properties of the Beck Depression Inventory-II, the Montgomery and Åsberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a sample from a healthy population. *Scandinavian Journal of Psychology*, *55*(1), 83–89. https://doi.org/10.1111/sjop.12090

Kleftaras, G., & Psarra, E. (2012). Meaning in life, psychological well-being and depressive symptomatology: A comparative study. *Psychology*, 3(4), 337–345. https://doi.org/10.4236/psych.2012.34048

Knafl, K., Deatrick, J., Gallo, A., Holcombe, G., Bakitas, M., Dixon, J., & Grey, M. (2007). The analysis and interpretation of cognitive interviews for instrument development. *Research in Nursing & Health, 30*(2), 224–234. https://doi.org/10.1002/nur.2019.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology, 50*, 537-567.

Lauritzen, S. L. (1996) *Graphical Models.* Oxford University Press, Oxford.

Leamy, M., Bird, V., Le Boutillier, C., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *The British Journal of Psychiatry: the Journal of Mental Science, 199*(6), 445–452. https://doi.org/10.1192/bjp.bp.110.083733

Leontiev, D. A. (2013). Personal meaning: A challenge for psychology. *The Journal of Positive Psychology, 8*(6), 459–470. https://doi.org/10.1080/17439760.2013.830767

Li, J.-B., Salcuni, S., & Delvecchio, E. (2019). Meaning in life, self-control and psychological distress among adolescents: A cross-national study. *Psychiatry Research, 272*, 122–129. https://doi.org/10.1016/j.psychres.2018.12.033

Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12*(4), 660–664. https://doi-org.ezproxy01.rhul.ac.uk/10.1177/1745691616687745

Lovibond, P.F., & Lovibond, S.H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scale with the Beck Depression Inventory and Anxiety Inventories. *Behaviour Research and Therapy, 33*, 335–343.

Marco, J. H., Pérez, S., & García-Alandete, J. (2016). Meaning in life buffers the association between risk factors for suicide and hopelessness in participants with mental disorders. *Journal of Clinical Psychology,* 72(7), 689–700. https://doi.org/10.1002/jclp.22285

McKenna, S. P., & Heaney, A. (2021) Setting and maintaining standards for patient-reported outcome measures: Can we rely on the COSMIN checklists? *Journal of Medical Economics, 24*(1), 502–511.https://doi.org/10.1080/13696998.2021.1907092

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. doi:10.1007/s11136-017-1765-4

Martela, F., & Steger, M. F. (2016). The three meanings of meaning in life: Distinguishing coherence, purpose, and significance. *The Journal of Positive Psychology, 11*(5), 531–545. https://doi.org/10.1080/17439760.2015.1137623

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation*, *27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

National Collaborating Centre for Mental Health. (2018). *The NHS Talking Therapies Manual* V6 [PDF]. https://www.england.nhs.uk/wp-content/uploads/2018/06/the-nhs-talking-therapies-manual-v6.pdf

Open Science Collaboration (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716. https://doi.org/10.1126/science.aac4716

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: The full life versus the empty life. *Journal of Happiness Studies, 61*(1), 25–41. https://doi.org/10.1007/s10902004-1278-z

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of life research: an international journal of quality of life aspects of treatment, care and rehabilitation, 27*(5), 1147–1157. https://doi.org/10.1007/s11136-018-1798-3

Salsman, J. M., Schalet, B. D., Park, C. L., George, L., Steger, M. F., Hahn, E. A., Snyder, M. A., & Cella, D. (2020). Assessing meaning & purpose in life: Development and validation of an item bank and short forms for the NIH PROMIS®. *Quality of Life Research*, *29*(8), 2299–2310. https://doi.org/10.1007/s11136-020-02489-3

Shanks, V., Williams, J., Leamy, M., Bird, V. J., Le Boutillier, C., & Slade, M. (2013). Measures of personal recovery: a systematic review. *Psychiatric services,* 64(10), 974–980. https://doi.org/10.1176/appi.ps.005012012

Schmittmann, V. D., Cramer, A. O. J., Waldorp, L. J., Epskamp, S., Kievit, R. A., & Borsboom, D. (2013). Deconstructing the construct: A network perspective on psychological phenomena. *New Ideas in Psychology*, *31*(1), 43–53. https://doi.org/10.1016/j.newideapsych.2011.02.007

Schulenberg, S.E., Schnetzer, L.W., & Buchanan, E.M. (2011). The Purpose in Life Test-Short Form: Development and psychometric support. *Journal of Happiness Studies, 12*, 861-876. https://doi.org/10.1007/s10902-010-9218-2

Solís Salazar, M. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, *27*(2), 192–200. https://doi.org/10.7334/psicothema2014.266

Steger, M. F. (2012). *Experiencing meaning in life: Optimal functioning at the nexus of well-being, psychopathology, and spirituality.* In P. T. P. Wong (Ed.), The human quest for meaning: Theories, research, and applications (pp. 165–184). Routledge/Taylor & Francis Group

Steger, M. F., Frazier, P., Oishi, S., & Kaler, M. (2006). The meaning in life questionnaire: Assessing the presence of and search for meaning in life. *Journal of Counseling Psychology, 53*(1), 80–93. https://doi.org/10.1037/0022-0167.53.1.80

Singh, K., Junnarkar, M., Jaswal, S., & Kaur, J. (2016). Validation of Meaning in Life Questionnaire in Hindi (MLQ-H). *Mental Health, Religion & Culture, 19*(5), 448–458. https://doi.org/10.1080/13674676.2016.1189759

Slade, M. (2009) *Personal recovery and mental illness: a guide for mental health professionals.* Cambridge: Cambridge University Press.

Slade, M., Amering, M., Farkas, M., Hamilton, B., O'Hagan, M., Panther, G., Perkins, R., Shepherd, G., Tse, S., & Whitley, R. (2014). Uses and abuses of recovery: Implementing recovery-oriented practices in mental health systems. *World Psychiatry, 13*(1), 12–20. https://doi.org/10.1002/wps.20084

Steger, F. M., & Samman, E. (2012). Assessing meaning in life on an international scale. *International Journal of Wellbeing, 2*(3), 182-195. https://doi.org/10.5502/ijw.v2.i3.2

Steinmann, I., Strietholt, R., & Braeken, J. (2022). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods, 27*(4), 667–702. https://doi.org/10.1037/met0000392

Revelle, W. (2021). psych: *Procedures for psychological, psychometric, and personality research* (2.1.9) [Computer software]. Retrieved from https://CRAN.R-project.org/package=psych.

Raykov, T. & Marcoulides, G. A. (2016). On the relationship between classical test theory and item response theory. *Educational and Psychological Measurement, 76*(2), 325–338. https://doi.org/10.1177/0013164415576958

Rosseel, Y., Jorgensen, T. D., Rockwood, N., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Scharf, F., & Du, H. (2022). lavaan: Latent variable analysis (0.6–10) [Computer software]. Retrieved from https://CRAN.R-project.org/package=lavaan.

Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology, 57*(6),1069–1081.

Ryff C. D. (2014). Psychological well-being revisited: advances in the science and practice of eudaimonia. *Psychotherapy and psychosomatics*, *83*(1), 10–28. https://doi.org/10.1159/000353263

Tait, L., & Lester, H. (2005). Encouraging user involvement in mental health services. *Advances in Psychiatric Treatment, 11*(3), 168-175. doi:10.1192/apt.11.3.168

Uher, J. (2021). Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *Journal of Theoretical and Philosophical Psychology, 41*(1), 58–84. https://doi.org/10.1037/teo0000176

Uher, J. (2022). Rating scales institutionalise a network of logical errors and conceptual problems in research practices: A rigorous analysis showing ways to tackle psychology's crises. *Frontiers in Psychology, 13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.1009893

Vallacher, R.R., & Wegner, D.M. (1987). What do people think they're doing? Action identification and human behaviour. *Psychological Review, 94,* 3–15. https://doi.org/10.1037/0033-295X.94.1.3.

van Bruggen, V., ten Klooster, P.M., van der Aa, N., Smith, A.J.M., Westerhof, G.J. and Glas, G. (2018), Structural Validity of the World Assumption Scale. *Journal of Traumatic Stress, 31,* 816-825. https://doi-org.ezproxy01.rhul.ac.uk/10.1002/jts.22348

van Weeghel, J., van Zelst, C., Boertien, D., & Hasson-Ohayon, I. (2019). Conceptualizations, assessments, and implications of personal recovery in mental illness: A scoping review of systematic reviews and meta-analyses. *Psychiatric Rehabilitation Journal, 42*(2), 169–181. https://doi.org/10.1037/prj0000356

Wang, Y.-H., & Liao, H.-C. (2015). Construction and validation of a meaning in life scale in the Taiwanese cultural context. *Psychological Reports, 117*(2), 437–451. https://doi.org/10.2466/08.07.PR0.117c19z4

Watkins E. (2011). Dysregulation in level of goal and action identification across psychological disorders. *Clinical psychology review*, *31*(2), 260–278. https://doi.org/10.1016/j.cpr.2010.05.004

Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, *44*(3), 219–246. https://doi.org/10.1177/0095798418771807

Weisz, J. R., Doss, A. J., & Hawley, K. M. (2005). Youth psychotherapy outcome research: A review and critique of the evidence base. *Annual Review of Psychology*, *56*, 337–363.

Welch, S., Klassen, C., Borisova, O., & Clothier, H. (2013). The DSM-5 controversies: How should psychologists respond? *Canadian Psychology, 54*(3), 166–175. https://doi.org/10.1037/a0033841

Wu, Y., Levis, B., Sun, Y., He, C., Krishnan, A., Neupane, D., Bhandari, P. M., Negeri, Z., Benedetti, A., & Thombs, B. D. (2021). Accuracy of the Hospital Anxiety and Depression Scale Depression subscale (HADS-D) to screen for major depression: Systematic review and individual participant data meta-analysis. *BMJ*, *373*, n972. https://doi.org/10.1136/bmj.n972

Xia, H.-Z., Gao, L., Wang, Y., Song, H., & Shi, B.-X. (2017). Development of the Chinese version of Meaning in Life Scale for cancer patients and psychometric evaluation. *Journal of Clinical Nursing*, *26*(21–22), 3298–3304. https://doi.org/10.1111/jocn.13675

Zhang, H., Sang, Z., Chen, C., Zhu, J., & Deng, W. (2018). Need for Meaning, Meaning Confusion, Meaning Anxiety, and Meaning Avoidance: Additional Dimensions of Meaning in Life. *Journal of Happiness Studies*, *19*(1), 191–212. https://doi.org/10.1007/s10902-016-9815-0

Zhou, J.-J., Tong, P., Ren, Q.-Z., Li, T., Zheng, Y.-J., Shen, Q.-Q., Liang, Y.-Y., & Gao, Y.-L. (2021). Development and validation of the Sources of Meaning in Life Scale for the Elderly in China. *Aging & Mental Health*, *25*(11), 2068–2077. https://doi.org/10.1080/13607863.2020.1792411

Appendix A

Measure of Mundane Meaning

INSTRUCTIONS: Below is a list of statements that someone might make about themselves and their life experiences. Please read each statement and decide how true the statement is of you *right now*. Then, choose a response corresponding to how true the statement is of you. Try not to think too much about each item--people are different, so there is no best answer.

EXAMPLE	NOT AT ALL TRUE OF ME						COMPLETELY TRUE OF ME
You can't always get what you want in life.	0	20	40	50	60	(80)	100

In the example, the number "80" has been circled, indicating that the statement is very true of the person responding, but not completely true.

	NOT AT ALL TRUE OF ME						COMPLETELY TRUE OF ME
1. I can picture what my life might be like far into the future.	0	20	40	50	60	80	100
2. Other people seem surer than me of where they are going in life.	0	20	40	50	60	80	100
3. I have developed new ways of looking at myself through my life experiences.	0	20	40	50	60	80	100
4. I feel like I have a mission in life.	0	20	40	50	60	80	100
5. I'm not certain that my life will amount to anything.	0	20	40	50	60	80	100
6. I can trust my intuition about how to handle ordinary affairs.	0	20	40	50	60	80	100
7. I know what's important in my daily life.	0	20	40	50	60	80	100
8. I have been able to find benefit from even my negative experiences.	0	20	40	50	60	80	100
9. I feel that my life is going somewhere.	0	20	40	50	60	80	100
10. The life I am now leading is not the one I was meant to lead.	0	20	40	50	60	80	100
11. I have lost sight of my goals.	0	20	40	50	60	80	100
12. I don't know what to expect from day to day.	0	20	40	50	60	80	100
13. I have confidence when dealing with everyday matters.	0	20	40	50	60	80	100

	NOT AT ALL TRUE OF ME						COMPLETELY TRUE OF ME
14. There is a feeling of coherence to my life.	0	20	40	50	60	80	100
15. My life has a sense of continuity.	0	20	40	50	60	80	100
16. I've lost the "thread" that used to run through my life.	0	20	40	50	60	80	100
17. I am living up to my potential.	0	20	40	50	60	80	100
18. The story of my life is unfolding in a satisfying way.	0	20	40	50	60	80	100
19. I have trouble feeling a part of my everyday roles.	0	20	40	50	60	80	100
20. I often find myself at a loss for what to do next.	0	20	40	50	60	80	100
21. Somehow my life has gone off track.	0	20	40	50	60	80	100
22. I don't have to think very hard about what I need to do from moment to moment.	0	20	40	50	60	80	100
23. I see a clear path forward for myself into the future.	0	20	40	50	60	80	100
24. I feel like I am in limbo.	0	20	40	50	60	80	100
25. I have been able to fit all my life experiences into my life story.	0	20	40	50	60	80	100
26. I have been able to make sense of difficulties that I have experienced in my life.	0	20	40	50	60	80	100
27. I need to stop and think before doing even ordinary things.	0	20	40	50	60	80	100
28. My sense of who I am in life is clear.	0	20	40	50	60	80	100
29. I have come to terms with events that have happened to me in my life.	0	20	40	50	60	80	100
30. I have a strong sense of purpose.	0	20	40	50	60	80	100
31. I can't trust my instincts in everyday matters.	0	20	40	50	60	80	100
32. Once I get up in the morning, I already have an idea of what I intend to do that day.	0	20	40	50	60	80	100
33. I have been able to put the past behind me and move on in my daily life.	0	20	40	50	60	80	100
34. I have a definite idea of my day-to-day priorities.	0	20	40	50	60	80	100
35. I'm not sure how the parts of my life fit together.	0	20	40	50	60	80	100