Measuring subjective well-being in later life: a review

ABSTRACT: This working paper assesses self-reported measures of subjective well-being in later life. In the first place, an overview of the theoretical background of a number of measures, focusing on those present in the English Longitudinal Study of Ageing (ELSA), is given. Secondly, the structure of these measurements and the interrelations between them are tested using confirmatory factor analysis. Thirdly, the cross-cultural measurement equivalence of the CASP-scale, a eudaimonic measure developed specifically for older adults, is testing using the Survey of Health, Ageing and Retirement in Europe (SHARE). These analyses reveal that it makes sense to distinguish affective, cognitive and eudaimonic measures of well-being empirically, but that these measures are more closely interrelated than one would expect on the base of theory alone. The analysis on CASP in SHARE reveals that the scale can be used to investigate differences in eudaimonic and hedonic subjective well-being across Europe, as partial scalar measurement equivalence is confirmed.

Bram Vanhoutte, CCSR Manchester.

www.ccsr.ac.uk

Introduction: Why measure well-being?

In the last decades, well-being has received increasing attention from both social scientists and government officials. On an international level, the OECD has considered measuring societal progress through objective indicators, such as the GDP, since its conception, but has included subjective measures in its statistics since the declaration of Istanbul in 2007. Similarly the EU Commission and Eurostat have launched initiatives to capture subjective components of well-being (Beyond GDP Conference in 2007). These developments on the international level have incited national and regional initiatives, among which the most influential are the 2009 French Commission on the Measurement of Economic Performance and Social Progress, headed by Joseph Stiglitz, Amartya Sen and Jean-Paul Fitoussi, and the more recent effort of the UK Office for National Statistics to Measure Well-Being (Beaumont, 2011) .

Although measuring subjective well-being is framed as a novel way to use social indicators to inform better policies, critics have pointed out that this is a very normative and individualistic way to look at societies problems, and that it tends to reinforce rather than overcome class barriers (Furedi, 2004; Lasch, 1979). The imperative to 'be happy', and the involvement of the state with one's emotional state, transfers the control over well-being to the hands of experts and therapists, disempowering the individual. This is paradoxically done under the moral disguise of the all importance of the self and the individual, and a symptom of what has been called our therapeutic age (Furedi, 2004; Lasch, 1979; Nolan, 1998; Szasz, 1999). The argument that the state should not try to influence individual subjective well-being, is echoed by proponents of the free market, who emphasize that GDP and employment are robust predictors of well-being, and the subjective aspect of it should be left to the individual to pursue (Booth, 2012).

The fairly recent policy interest in measuring subjective well-being is based on a longer tradition of academic research into quality of life (Nussbaum & Sen, 1993) and positive psychology (Seligman & Csikszentmihalyi, 2000), aimed at extending the focus of research in the behavioural sciences from problematic behaviour to positive qualities, from repairing and healing to enhancing the ability ofindividuals to maintain a good life (Seligman & Csikszentmihalyi, 2000). In the framework of the ageing of the population, it can be said that measuring subjective well-being and enhancing a good later life are even more important. As people are living longer, and are spending a significant part of their later life in good health, a new demographic category, labelled the third age, has emerged (Laslett, 1989). This structural change at the level of the population translates itself into a new life stage for the individual as well. As the responsibilities of employment and childcare fade away, this life phase creates the possibility to fulfil personal life goals and dreams, given good health and relative wealth. As illness and other problems associated with age set in, the fourth age, secluded from society and increasingly dependant on others, starts as a final life phase. The third age perspective has received severe criticisms, with claims that it is a middle class perspective on retirement and doesn't incorporate any reference to social inequalities (Bury, 1995).

In this briefing paper an overview of the existing approaches to examine subjective well-being in later life is given, based on available measures. We will focus on the subjective measures of well-being, but acknowledge that different approaches such as objective lists of conditions from which well-being emerges (Nussbaum & Sen, 1993) or preference satisfaction (Dolan & Peasgood, 2008) also have their merits. Both theoretical background and methodological issues of the measures are

addressed. An important division in measuring instruments is made on the basis of different philosophical backgrounds of what well-being actually entails (Ryan & Deci, 2001). Is subjective well-being mainly about being happy, or are there other things than pleasure and pain, such as self-actualisation, that influence one's level of contentment? These different approaches to well-being, classified as respectively hedonic and eudaimonic measures, will be a first point of attention. A second point of attention is to evaluate how scales that capture different aspects of well-being look when applied to the English Longitudinal Study of Ageing (ELSA). Do the structural models mentioned in the literature, usually tested on either relatively small samples of university students or large scale population surveys, also fit people aged 50 or older in England? We evaluate the scales by examining the interrelations between different scales, so that we can assess to what extent they differ from each other. In a final step measurement equivalence of the CASP scale (Hyde, Wiggins, Higgs, & Blane, 2003) across different cultures will be investigated using the Survey of Health, Ageing and Retirement in Europe (SHARE). Cross-cultural measurement equivalence means that the scale captures the same concept in different countries, and that scores on the scale can be compared.

1. Different approaches to measuring subjective well-being

Although in everyday life subjective well-being (SWB) is probed for by the straightforward question "How are you?", accurate and reliable assessment of well-being is at the base of a quite complex and substantial debate. A first point that needs to be addressed is what subjective well-being actually entails.

Subjective well-being is often used in conjunction with physical health, and is commonly used as a concept for psychological health. Secondly, it is seen as the subjective counterpart of objective indicators for quality of life, and involves an individual judgement. A third point which defines subjective well-being, is that, just like it's counterparts madness and illness, it is at least partly a social construct. What wellbeing entails therefore depends not only on the psychological outlook one has on life, but equally on the position in society and the society one lives in. This makes any enquiry into the nature of well-being a meeting ground between philosophical theory and empirical measurement (Sumner, 1999).

1.1 Hedonic well-being

The hedonic view on well-being assumes that through maximizing pleasurable experiences, and minimizing suffering, the highest levels of well-being can be achieved. This emphasis on pleasure and stimulation entails not only bodily or physical pleasures, but allows any pursuit of goals or valued outcomes to lead to happiness. Both cognitive and affective aspects of well-being can be identified within this approach (Diener, 1984). A high level of well-being in the hedonic approach consists of a high life satisfaction, the presence of positive affect and the absence of negative affect (Diener, 1984). Well-being resides within the individual (Campbell, Converse, & Rodgers, 1976), and therefore does not include reference to objective realities of life, such as health, income, social relations or functioning.

The affective aspect of hedonic well-being consists of moods and emotions, both positive and negative. Positive and negative affect each form a separate domain, and are not just opposites (D. Watson, Clark, & Tellegen, 1988). Positive affect (PA) is a state wherein an individual feels enthusiastic, active and alert. High PA means high energy, full concentration and pleasurable engagement, while low PA encompasses sadness and lethargy. Negative affect generally captures subjective distress and unpleasurable mood states, such as anger, disgust, guilt, fear and nervousness. Low NA on the other hand encompasses calmness and serenity. Both positive and negative affect are usually measured by letting the respondent assess the prevalence of a number of emotional states in the last month (D. Watson et al., 1988). The affective approach to well-being can be traced back to the first enquiries on psychological well-being and quality of life (Bradburn, 1969).

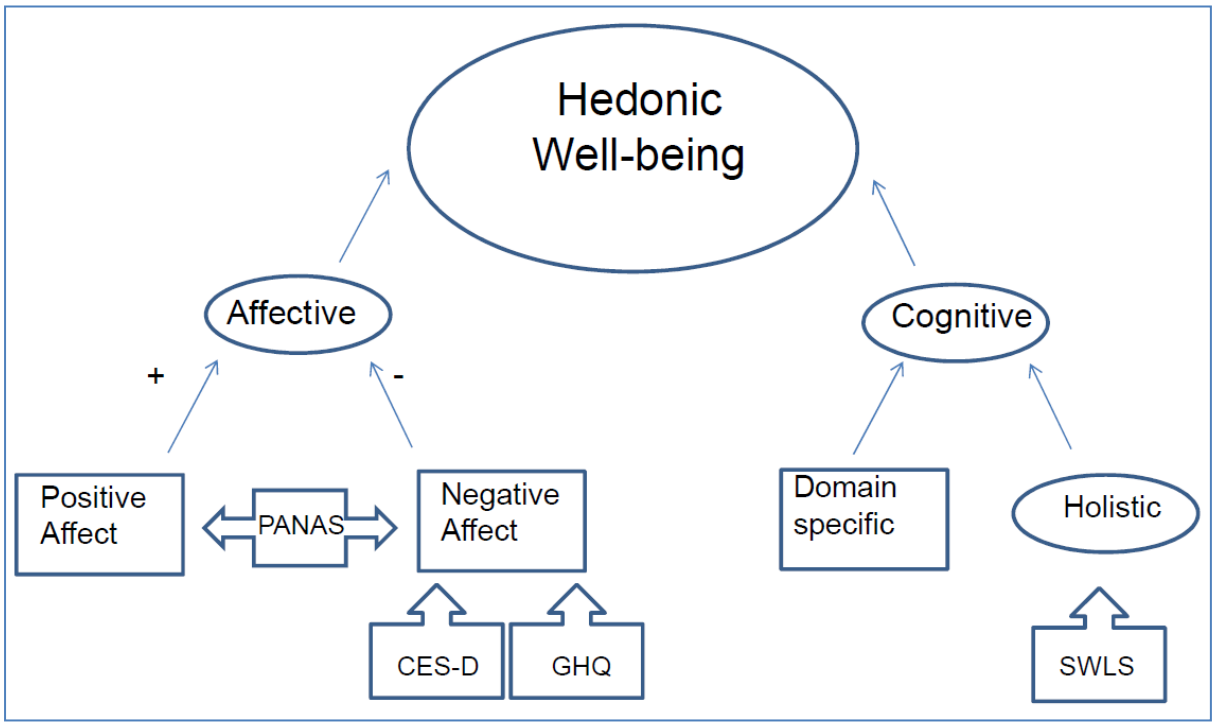
The affective aspect of well-being brings measurement very close to assessing mental health. Therefore it is not surprising that depressive symptoms are sometimes used as a measure of NA (Demakakos, McMunn, & Steptoe, 2010). Depression is traditionally assessed by the CES-D scale (Radloff, 1977), which has been shown to be accurate and valid among the older population as well as at younger ages (Lewinsohn, Seeley, Roberts, & Allen, 1997). A second measure for mental health, the 12 item version of the General Health Questionnaire (GHQ) (Goldberg, 1988) can be seen in the light of affective measures of SWB as well. The GHQ-12 is a widely used screening tool for psychiatric disturbance, and has shown to have good psychometric properties and reliability for older people (Y. B. Cheung, 2002).

In relation to later life, affective aspects of well-being have been studied quite intensively. On the level of measurement, it has been illustrated that the PANAS scale (D. Watson et al., 1988) has good psychometric and scale properties among the old, and yields information that is comparable to other age groups (Crawford & Henry, 2004; Kercher, 1992; Kunzmann, Little, & Smith, 2000). In regard to differences in mean levels of affect, it is an established fact that NA decreases over the lifespan, albeit the rate of decline is slower in old age, and may reverse in old-old age, while results for PA are not unequivocal (Charles, Reynolds, & Gatz, 2001; Crawford & Henry, 2004; Kunzmann, 2008; Kunzmann et al., 2000; Ready et al., 2011). On the level of facets of emotions, there is some evidence that although PA and NA are valid and separate factors, the structure of the interrelations among emotions in older adults differs from younger adults (Ready et al., 2011). Specifically sadness and depressive feelings seem to be more interrelated with anxiety. In connection to that, some studies report more somatic symptoms than emotional moods of depression by older adults (King & Markus, 2000), leading to the challenged idea that depression manifests itself in a different way for older adults, a phenomenon called later life depression (Alexopoulos, 2005; Parmelee, 2007). As depression is not a monolithic disease, but an emotional disorder accompanied by physiological symptoms, it is difficult to distinguish it from conditions in later life that trigger similar symptoms, such as chronic illness or cognitive impairment as the result of dementia or Alzheimer's disease (Parmelee, 2007). In addressing this issue, it is helpful to make a distinction between major depression, which is less prevalent among the elderly (2%), and minor depression (15%), which is more common, and closely interrelated with stressful life events in later life and vascular risk factors (Beekman & Deeg, 1995; Van den Berg et al., 2001). While the CES-D scale and GHQ have been shown to be a robust measurement of major depression in later life, they show to be less accurate in picking up minor depression (Papassotiropoulos, Heun, & Maier, 1999; L. C. Watson & Pignone, 2003).

The cognitive component of hedonic well-being, often referred to as life satisfaction, is a judgemental process in which individuals asses the quality of their life based on their own set of criteria (Pavot & Diener, 1993). As such, it differs from domain specific evaluations of satisfaction (Campbell, Converse, & Rodgers, 1976) in that an idiosyncratic set of standards is taken into account, which allows for comparing satisfaction with life over groups of people with different aspirations in life. The Satisfaction With Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985; Pavot & Diener, 1993) consists of 5 Likert items to be rated on a response scale ranging from 1 (strongly disagree) to 7 (strongly agree), inviting respondents to make a global evaluation of their life. It was also explicitly tested on older respondents (Diener et al., 1985). From a methodological perspective, it is surprising that all the items are worded in a positive way, because this way the scale could suffer from extreme response and acquiescence bias.

Critics Perceptions about the self and one's own life tend to be too positive and optimistic (Kahneman & Thaler, 2006; Taylor & Brown, 1988), so that hedonic well-being ultimately depends on how high or low one sets his goals. This judgemental relativity is seen as a major problem in assessing the validity across the population for hedonic cognitive measures, as even a slave can be happy. Similarly, adaptation plays a main role in the cognitive process of accepting the circumstances as they are and moving to a normal level of well-being (see further). A second severe criticism on well-being as maximizing pleasure, is that negative events have an important role in providing insight about one-self, or growing as a person (Ryff & Singer, 1998). Positive psychology itself is deeply rooted in investigating which type of persons are resilient to negative conditions (Seligman & Csikszentmihalyi, 2000).

Figure 1: schematic representation of measures of hedonic well-being



1.2 Eudaimonic well-being

A second, and in practice largely complementary (Waterman, 1993), approach to well-being starts from a different concept of well-being. A good life is not just about pleasure and happiness, but involves developing one-self and realizing one's potential (Ryff & Keyes, 1995). Eudaimonic well-being reflects positive functioning and personal expressiveness. Positive functioning, or psychological well-being, reflects the need for self-actualisation in Maslow's (1968) need hierarchy. Similarly, positive functioning can be seen from the perspective of developmental psychology, as personality changes articulate well-being as trajectories of continued growth across the life cycle (Erikson, 1959).

As the concept of positive functioning is rooted in different approaches, several different measurement instruments can be found. Ryan and Deci (2000) conceptualize it in their self-determination theory and see autonomy, competence and relatedness as three basic necessities for personal growth, integrity and well-being. By looking at six distinct aspects of actualisation (autonomy, personal growth, self-acceptance, life purpose, mastery and positive relatedness), Ryff & Keyes (1995) measure psychological well-being, which they see separate from subjective well-being.

In the framework of studies on later life, a measure specifically targeted at older populations has been developed (Hyde et al., 2003). Four constructs, namely Control, Autonomy, Self-realization and Pleasure (CASP) together can be seen as an accurate measure of positive functioning, and subjective quality of life in later life. An explicit aim of this measure from it's conception was to distinguish quality of life from it's drivers, such as health (Hyde et al., 2003). Therefore it is quite surprising to see explicit references to the respondents' age and health on the item level, in items such as "My age prevents me from doing the things I would like to" and "My health stops me from doing the things I want to do". Theoretically this is unsound because it contaminates the measure with aspects of health status. From a methodological point of view, a confirmatory factor analysis by the developers of the measure has equally shown that the error term of the item referring to health correlates with some other items in the scale, and that the scale shows better properties in a reduced form with 12 items (Wiggins, Netuveli, Hyde, Higgs, & Blane, 2007). A second point, that is of importance for this study concerns the domain of Pleasure, which could be seen more as a hedonic than a eudaimonic form of well-being. When looking at different measures of well-being at the same time, this should be kept in mind.

Comparing the dimensionality of different conceptualisations of eudaimonic well-being it becomes clear that in large lines they rely on very similar concepts and sub-dimensions (Table 1). All three approaches depart from the idea that human flourishing depends on the satisfaction of certain psychological needs. Autonomy is a need that is present explicitly in psychological well-being (PWB), self determination theory (SDT) and CASP. Both control in CASP, and environmental mastery in PWB can be seen as a closely related concept, relating to autonomy. The second key aspect of eudaimonic well-being is developing one-self, and is captured as personal growth in PWB, as competence in SDT and self-realisation in CASP. The largest difference between the three approaches is that both PWB and SDT do not see pleasure, or any other aspect of Diener's hedonic subjective well-being concepts as an explicit psychological need (Diener, Sapyta, & Suh, 1998; Ryff & Singer, 1998), while CASP does. While Ryff & Singer (1998) downplay the importance of subjective well-being altogether, Ryan & Deci (2001) see it as a consequence of the fulfilment of needs, that goes hand in hand with

eudaimonic well-being. Secondly, relatedness, or having warm and positive social relations, is seen as an essential need for psychological wellbeing, while it is not explicitly defined in the CASP scale.

Table 1: Overview of dimensions of eudaimonic well-being

PWB (Ryff & Keyes, 1987)	**SDT (Ryan & Deci, 2000)**	**CASP 19 (Hyde et al. 2003)**
Autonomy	Autonomy	Autonomy
Personal Growth	Competence	Self-realisation
Self-acceptance		
Life Purpose		
Environmental mastery		Control
Positive Relatedness	Relatedness	
		Pleasure

1.3 Retrospective, Experienced and Reconstructed Well-being

A second form of measurement diversity reflects both theoretical and methodological considerations on the nature of changes in well-being. Is well-being a relatively stable stock product, affected little by fluctuations over time and life-events, or can it better be characterised as a flow, volatile and changeable? In the context of well-being in later life, the evolution of well-being over time is specifically interesting, as old age is often characterised as a period in life where health risks and social losses occur simultaneously or within a short time-span.

One way to look at well-being is to see it as experienced utility in the classical economical sense. Probing for someone's level of well-being as a stock, by using self reporting in surveys, can be prone to errors because of effects of social desirability judgement and memory, which have been illustrated extensively in the case of hedonic well-being (Kahneman & Thaler, 2006). Nevertheless, research has shown that both hedonic and eudaimonic self-reported well-being to be closely associated to the attribution of positive personality traits by both acquaintances and clinicians, and cheerful, socially skilled behaviour, which illustrates that self-reports are grounded in reality (Kahneman & Krueger, 2006; Nave, Sherman, & Funder, 2008).

To emphasize the flow of hedonic well-being, alternative methods of collecting information have been set up. One influential but time-consuming approach is experience sampling (Csikszentmihalyi, 1990), where people report their moods and emotions on the spot in everyday life, by describing the activity they are doing and the pleasure achieved from it when a timer beeps, which happen several times during a day. In a recent effort to make this information easier to acquire, the day reconstruction method, where the respondent reconstructs his previous day episode by episode and then assigns moods to each period, has shown to be a reliable equivalent (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004).

A different approach to changes in well-being focuses on the impact of positive and negative effects of life events and changes in conditions. The main question focuses on the treadmill effect, meaning that well-being levels adapt to both positive and negative events and emotions, so that there is no actual evolution in the long term (Brickman & Campbell, 1971; Diener, Lucas, & Scollon, 2006). Although there initially was substantive evidence for the treadmill effect when looking at hedonic measures of well-being (Brickman, Coates, & Janoff-Bulman, 1978), some substantial revisions to the treadmill argument have been suggested (Diener et al., 2006). A first domain of concern is the so called set points – the levels of well-being that one departs or returns from when experiencing an event. These points are multidimensional, meaning that they can differ for affective and cognitive aspects of well-being. Set point also are not neutral, but instead tend to be positive (Diener & Diener, 1996), and vary considerably among individuals, due to inborn personality based influences (Diener, Suh, & Lucas, 1999). Secondly, while the treadmill argument implies that people eventually adapt the both good and bad circumstances, it has been illustrated that change does happen on the long term, for example when faced with unemployment (Lucas, Clark, Georgellis, & Diener, 2004), or loss of a partner (Lucas, Clark, Georgellis, & Diener, 2003). The extent to which adaptation occurs is heavily dependent on the individual as well, and coping and personality characteristics seem to play an important role. It has to be kept in mind that the bulk of the research on this topic has examined hedonic well-being. Nonetheless, also when it comes to eudaimonic well-being processes of adaptation can be thought of, especially when looking at self-realisation (Waterman, 2007). The

experience of flow (Csikszentmihalyi, 1990), when the challenge posed and the skill of an individual are balanced, could become quite rare as a person is becoming more experienced and hence more skilled, leading to an eudaimonic treadmill. Waterman (2007) argues that the opposite is actually the case, since eudaimonic well-being is the result of striving more than the actual outcome, and new fields for self-realisation are in pratice endless.

In this analysis we will limit ourselves to the traditional self-reported measurements of hedonic and eudaimonic well-being, but it is clear that alternative measures are possible and available.

Assessing measurement

The measurement instruments of well-being mentioned and present in ELSA will be investigated in more detail in this analysis. While some scales were specifically designed for on older population (CASP), others are scales (SWLS, CES-D, GHQ) usually applied to a general population sample. Therefore it is important to look at the structure of these scales specifically for an older population, and to look if they measure different concepts of well-being in the same way as they do in the general population. Since CASP is a relatively novel, specific and complex measure, and the only measure in ELSA for the eudaimonic aspects of wellbeing, we will treat it in greater detail.

It is beyond the scope of this paper to examine all possible aspects of the measurement of well-being. In this analysis we limit ourselves to two points. First, what is the structure of the different scales? This research question gives insight into the theoretical nature of well-being: Can well-being be seen as a single dimension or not? To what extent to different scales reflect different aspects of well-being? The best way to test this, is to first identify the ideal structure for the different aspects of subjective well-being, reflected in different scales. In a next step, a second-order model of well-being is constructed, by looking if and how the different sub-dimensions relate to each other. A second point of attention is the measurement of well-being over different subgroups. All too often a measurement instrument is used to compare groups, without investigating if the instrument functions in a similar way across groups. In this paper, the measurement invariance across European countries of the CASP scale will be investigated.

The first research question, on the structure of subjective well-being, will be investigated using the first three waves (collected in respectively 2002, 2004 and 2006) of the English Longitudinal Study of Ageing (ELSA) (Marmot et al., 2011)[1]. Different waves were used, because although not all instruments were present in the first or second wave, they have larger sample sizes (respectively 10253 and 8780) and as such allow for greater variability in the data. The third wave (using both core sample members and the refreshment sample, in total 8598 respondents) is used to asses the interrelations beween all available scales. More detailed descriptive statistics on the data used can be found in appendix.

The second research question, investigating the cross-cultural equivalence of CASP, will be examined using wave 2, collected in 2006/2007, of the Survey of Health, Ageing and Retirement in Europe (SHARE)(Börsch-Supan & Jürges, 2005)[2]. Wave 2 is used since more countries took part, which gives

[1] The data were made available through the UK Data Archive (UKDA). ELSA was developed by a team of researchers based at the National Centre for Social Research, University College London and the Institute for Fiscal Studies. The data were collected by the National Centre for Social Research. The funding is provided by the National Institute of Aging in the United States, and a consortium of UK government departments co-ordinated by the Office for National Statistics. The developers and funders of ELSA and the Archive do not bear any responsibility for the analyses or interpretations presented here.

[2] This paper uses data from SHARELIFE release 1, as of November 24th 2010 or SHARE release 2.5.0, as of May 24th 2011. The SHARE data collection has been primarily funded by the European Commission through the 5th framework programme (project QLK6-CT-2001- 00360 in the thematic programme Quality of Life), through the 6th framework programme (projects SHARE-I3, RII-CT- 2006-062193, COMPARE, CIT5-CT-2005-028857, and SHARELIFE, CIT4-CT-2006-028812) and through the 7th framework programme (SHARE-PREP, 211909 and SHARE-LEAP, 227822). Additional funding from the U.S. National Institute on Aging (U01 AG09740-13S2, P01 AG005842, P01 AG08291, P30 AG12815, Y1-AG-4553-01 and OGHA 04-064, IAG BSR06-11, R21 AG025169) as

us more variability (33657 respondents in 17 countries). More detailed descriptive statistics on the data used can be found in appendix.

An important aspect of the measurement of well-being is investigating the structure of scales commonly used. Factor analysis is a good tool to assess measurement adequacy. Two main forms of factor analysis can be distinguished: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA is more data-driven, and is often used in scale development, when there is little underlying theory on how items should load on a factor, or how many factors are present. CFA is used to test and confirm theoretical hypotheses on scale structure. As we are working with existing and widely used scales, which have substantive theoretical hypothesis attached to them, CFA will be used. A specific application of CFA is assessing measurement equivalence of instruments. To be sure that differences in scales between different (sub)populations reflect real differences, and are not measurement artefacts, a level of measurement equivalence is necessary. In the following part I will outline the different steps and the criteria for decision in each step in looking at a scale. I depart from the available measures in ELSA, and build on existing research. A last important note is that while this kind of analysis illustrates problems associated with measurement, it does not insinuate that analyses based on "bad" versions of a scale are flawed in themselves. Measurement models are very useful in testing the latent structure behind a scale, but usually a refined scale does not alter substantive analysis to a large extent. As such this analysis should be seen more of a test of the theoretical background of the concept of well-being.

Usually maximum likelihood estimation (MLE) is used to estimate CFA models, but although this method is more precise for parameter estimation, it's limited to estimating a small number of factors (2 or 3). We will use the weighted least squares means and variances adjusted (WLSMV) estimator, that is computationally more efficient and gives equally reliable estimates as MLE (Beauducel & Herzberg, 2006). A positive aspect of this method is that it does not assume normality of the distribution over the different answering categories. A drawback of this estimation method is that it gives less comparable information on model fit, because the chi-square based statistics cannot be directly compared between nested models as in MLE. This only becomes important in the next step of our analysis, when looking at measurement equivalence.

To determine which model fits better, a number of test statistics are available. We will focus on the most widely used ones, namely the Root Mean Square Error of Approximation (RMSEA) (lower than .8 for decent fit and lower than .06 for good fit), the Comparative Fit Index (CFI) (higher than .95 for good fit) and Tucker Lewis Index (TLI) (higher than .95 for good fit) (Hu & Bentler, 1999) . Similarly the size of factor loadings will be looked at, because the use as a sum scale requires all items to load equally good (more than .60) on the latent constructs. A low factor loading means that in practice the item does not contribute a great deal to the latent measure.

well as from various national sources is gratefully acknowledged (see www.share-project.org for a full list of funding institutions).

2.1 Identifying the best structural factor model

The first step in looking at the way in which a latent scale captures the variability present in separate items consists of making the best configuration of items and factors. The idea in this first step is to make the best possible model for the data based on substantive theory. Which items adequately define a scale? Especially when items are simply summed up, as is the case in CES-D, GHQ and CASP scales, it is of utmost importance that each item is defined by the latent concept similarly, and that there are no large differences in factor loadings. Another issue that is narrowly intertwined with the chosen items, is the number of factors, or sub dimensions that exist in a scale. In exploratory factor analysis, the data provides a certain number of dimensions and it's up to the researcher to determine the criterion for cut-off. The extensive use of EFA in making latent factors has been criticized, as it does not allow examination of measurement bias. This means that EFA assumes that variables are being perfectly measured, without any form of measurement error, and that all of an observed measure's variance is true score variance (Brown, 2006). It has been shown that a false number of factors can surface if method effects are not taken into account (Brown, 2003; Chen, Rendina-gobioff, & Dedrick, 2010; DiStefano & Motl, 2009; Hankins, 2008; Van de Velde, Bracke, Levecque, & Meuleman, 2010; Wood, Taylor, & Joseph, 2010). In particular, items posed in a negative manner can provoke different answering patterns of a respondent, that do not relate to the substantive matter of the scale but rather to the fact that the item is worded negatively (Marsh, 1996). In other words, asking someone 'how often are you unhappy', is not simply the inverse of 'how often are you happy'. To account for these effects, one can either make a separate uncorrelated method factor, on which negatively worded items load, or allow error correlations between negatively worded items in the scales. As such we will depart from different theoretical expectations on how the items fit together and identify the best model for the data.

1.1.1 CASP

The CASP scale in its original form has 19 items, but a revised form of 12 items has been proposed for use (Wiggins et al., 2007). It has been used in the self-completion questionnaire in the 19 item form in ELSA waves 1-5 and the 2004 wave of the US based Health and Retirement Study (HRS), and in a 12-item form in SHARE. The 12 item version of CASP used in SHARE is not same as the preferred 12-item version, as the choice of items was based on preliminary analysis. Since in a lot of analysis using the CASP scale the items themselves are not mentioned and only the sum scale is used, but people refer to psychometric tests on the original scale, it is quite important to investigate the structure of the latent concept in all versions of the scale. In the original study that tested the qualities of the CASP scale, a first order factor solution based on 4 sub dimensions was proposed for the 19 item scale, and a similar factor structure based on three sub dimensions was proposed for the 12 item version (Wiggins et al., 2007).

In our analysis we will replicate the confirmatory factor analysis of Wiggins et al. (2007), and see if a method effect accounting for the negative item wording significantly improves the fit of the model to the data, in our case the first wave of ELSA. Additionally, as we can theoretically expect a division between eudaimonic and hedonic aspects of well-being, a two factor solution isolating the domains pleasure from control, autonomy, and self-actualisation will also be tested. Understanding the differences between these models is key to grasping how confirmatory factor analysis will be used to test theoretical models, therefore a schematic representation of the models can found in the

appendix (figures A-E). The baseline model (figure A) assumes all items load onto the same factor (Figure 2). Each item is associated with an item-specific error term, which represents the variation that is not accounted for by the latent factor, in this case the CASP scale. To account for the possible measurement bias introduced by negative item wording, two possible specifications are used interchangeably in the literature. A first option is to allow correlations between the error terms of the items that are phrased negatively (figure B). A second option is to specify a latent factor onto which these items load, next to their loading onto the substantive factor (figure C). If less than three items are phrased in a different direction, the option with error correlations is more sensible as a latent factor needs at least three items to be identified. The number of dimensions of the latent factor is another question that needs to be addressed. In the case of the CASP scale, originally four sub-dimensions were proposed (Hyde et al., 2003; Wiggins et al., 2007). Specifying a high number of factors can lead to problematic results, as factors that are too similar to each other do not discriminate concepts enough to make empirical sense, which is indicated by a non positive definite covariances, and correlations higher than one between factors. As mentioned previously, not accounting for reverse item phrasing can also inflate the number of factors that surface, so it is important to test for the combination of a higher number of factors and at the same time account for this phrasing bias. It has already been shown that the domains autonomy and control are closely related (Wiggins et al., 2007), and can be seen as one latent sub-dimension, resulting in a three factor structure (figure E). Looking at the philosophical foundations of well-being, it can be hypothesized that pleasure in itself could be seen as a separate hedonic dimension, more focusing on enjoyment, while control, autonomy and self-actualisation are more eudaimonic, and related to freedom and goal realisation. This theoretical approach assumes a two factor solution (figure D). A last possible variation is to construct a second order factor, onto which each sub-dimension loads, as in the original CASP proposal. Trying to specify closely related concepts can lead to standardised factor loadings higher than one and negative covariance. Nevertheless, when concepts correlate highly, it can be safely assumed that they refer to the same latent dimension.

Table 2: CFA for CASP 19 in ELSA wave 1

	RMSEA	CFI	TLI	items with low standardised factor loadings (<.4)	
1 factor	.121	.848	.829	f, i	
with error corr	.103	.901	.877	f, i	
with method factor	.106	.889	.870	f, i	
2 factor	.115	.865	.847	f, i	
with error corr	.096	.915	.893	f, i	
with method factor	.100	.902	.885	f, i	
3 factor	.106	.887	.870	f, i	
with error corr	.096	.916	.893	f, i	
with method factor	.099	.905	.886	f, i	
4 factors	.107	.886	.866	f, i	Problem with f2
with error corr	.096	.919	.894	f, i	Problem with f2
with method factor	.099	.906	.885	f, i	Problem with f2
1 higher order factor with 3 subdimensions	.106	.887	.870	f, i	Problem with f3
with error corr	.096	.916	.893	f, i	Problem with f3
with method factor	.099	.905	.886	f, i	Problem with f3
1 higher order factor with 4 subdimensions	.116	.865	.844	f, i	
with error corr	.096	.917	.893	f, i	Problem with f1
with method factor	.101	.901	.881	f, i	Problem with f1

In Table 2 above, the fit statistics for the different models for the 19 item version of CASP are presented. We are partly replicating the analysis of the conceivers of the scale, investigating to what extent a method factor compares to their findings (Wiggins et al., 2007). Although the exact fit indexes could not be replicated (due to changes in the way Mplus estimates the models in different versions of the program (personal communication with L. Muthen)), similar conclusions can be drawn. In general terms, it can be said that CASP 19 does not perform very well to any of the proposed models in terms of model fit. In each model, including a method factor or allowing error correlations between negatively worded items improves the model fit, pointing to the importance of account for negative wording. In more complex models with a lot of factors, non positive definite covariances surface, which illustrates the frail nature and close relatedness of the different factors.[3] Both the higher order models and four factor model suffer from this severe limitation. The three factor model merges the first and second factor, respectively control and autonomy, as it seems problematic to try and separate them. For the two factor model, self-actualisation was added to the control/autonomy dimension, to test to what extent a simple split between eudaimonic and hedonic measures provides a better and more parsimonious model.

Some items (limiting effect of age and health, item a and h respectively) from the control and autonomy domains were loading better or equally well on the method factor as on the substantial

[3] These translate themselves in practice into correlations higher than 1 between factors, and is a sign that the discriminatory power of two dimensions is not high enough to see them as separate.

factor, pointing to the fact that these items are measuring something else than control or autonomy. In the case of these items, which grasp the extent to which age and health impose limitations, it is relatively unsurprising that they do not measure control and autonomy[4]. Both items i (shortage of money) and f (family responsibilities) have consistently low factor loadings, meaning that they do not adequately reflect the latent factors. Further investigation, using EFA in a CFA framework (Marsh et al., 2009), showed a number of cross-loadings. This means that items load strongly on a different factor than the one they are assigned to, for example item o on factor 1. Similarly, a quite different factor structure emerges from the data than that originally proposed by the authors of the scale when using EFA (see appendix). Therefore, if we want a parsimonious model including all of the items, there seems to be little reason to choose a four factor model over a three or even two factor one. Correlations between the factors are high or very high, ranging from .748 to .919 in three factor model, so that we safely can assume that the same latent concept, quality of life, is measured. In conclusion, we can say that the factor structure of the CASP 19 scale is rather problematic, and ideally a different structure should be proposed and/or some items should be deleted.

As such, it is quite justified that the proponents of the scale suggest a shorter scale, eliminating some of the problematic items. To assess to which extent this scale is an improvement, we do a similar analysis as on the shortened CASP scale proposed by the developers of the scale (Table 3), and the different 12 item scale included in SHARE. The shortened 12 item scale excludes items c, f, h, m, n, p and q. The SHARE version of the 12 item scale excludes items c, g, h, l, m, p and q.

Table 3: CFA for CASP 12 in ELSA wave 1

	RMSEA	CFI	TLI	items with low standardised factor loadings (<.4)
1 factor	.116	.915	.869	i
with error corr	.101	.942	.920	i
with method factor	.103	.936	.918	I
2 factor	.103	.933	.917	I
with error corr	.088	.987	.940	I
with method factor	.091	.951	.936	i
3 factors	.096	.944	.928	i
with error corr	.091	.957	.936	i
with method factor	.093	.951	.932	i
Second order factor	.096	.944	.928	i
With error corr	.091	.957	.936	i
With method factor	.089	.956	.939	i

The reduced scale in general has a better model fit. As some of the items that had a low factor loading are removed, we also run into fewer problems in terms of model specification. One item that has a low factor loading, which remains in the 12 item version, is the extent to which money plays a

[4] The concept of frailty could be a useful in this context. Frailty is seen as a clinical and biological syndrome of decreased reserve and resistance to stressors, causes vulnerability to adverse outcomes, and highly related to ageing (Fried et al., 2001)

role in autonomy (Factor 1), and as such can be seen as substantively interesting as an item on itself. It should not be seen as a good indicator of autonomy or control nonetheless. Again the inclusion of a method factor or error correlations for negatively worded items improve the fit. With the reduced scale, a two factor solution seems slightly better and is also more parsimonious than a three factor or second order solution. With correlations of .845 between the factors, we can again safely assume one latent dimension of eudaimonic SWB. Note that the correlations are markedly higher when controlling for the method factor.

Table 4: CFA for CASP 12 – Share version in Elsa wave 1

	RMSEA	CFI	TLI	items with low standardised factor loadings (<.4)
1 factor	.112	.912	.893	f, i
with error corr	.090	.954	.931	f, i
with method factor	.092	.947	.928	f, i
2 factor	.101	.930	.913	f, i
with error corr	.073	.970	.954	f, i
with method factor	.078	.963	.948	f, i
3 factors	.092	.944	.928	f, i
with error corr	.076	.970	.951	f, i
with method factor	.079	.962	.946	f, i
Second order factor*	.092	.944	.928	f, i
With error corr*	.076	.970	.951	f, i
With method factor*	.080	.962	.945	f, i

*Problem with f1

The results of the SHARE version of the 12 item scale are similar to the general 12 item version. Because of the different items included, there are some problems in the second order factor structure, that did not surface in the general 12 item version. The fit also seems especially good for both the two and three factor models that account for negative item wording, although this version includes some weakly loading items.

What can we conclude from the replication of the original analysis? On the one hand, it is clear that the CASP scale is best seen as a two dimensional scale, or at most comprising three dimensions rather than four. Control, autonomy and self-actualisation are too close in empirical terms to be defined as separate dimensions, especially when the number of items is more limited. A second conclusion is that there is an effect of negatively worded items on the total scale, which should be taken into account, regardless of the model that is being used. This method effect, if not taken into account, perturbs the score on the latent factors and as such results in lower correlations between the latent factors. A third conclusion is that not all items seem to be good indicators of the latent scale. Some have a low loading, while others show cross loadings or even higher loadings on the method factor than on the substantive factors.

Constructing a new robust and theoretically rigorous CASP scale

As such these findings suggest that the CASP scale, even the 12 item version, can be significantly improved, although this will be at the cost of the number of items and therefore will reduce the variance of the scale, which may limit its practical use in distinguishing different levels of quality of

life. Working with the available items, but adhering to a number of strict methodological rules, we propose a CASP scale that provides a more robust measurement of well-being. In our version of the scale, all standardized factor loadings have to be higher than .40, and also be higher on the substantial factor than on the method factor. Similarly we want the measure to be independent of possible drivers of quality of life, such as health and age, and strive for a balanced factor solution where possible, meaning that each dimension is captured by roughly the same amount of items.

For CASP19 this results in a 15 factor scale excluding items a, f, h and i. All these items refer to issues limiting the respondent in his or her freedom (respectively age, family, health and money), and as such are on the borderline between objective and subjective indicators. The items on family and money consistently had very low loadings, and are therefore not good indicators of either autonomy or control, or quality of life in general. The items on age and health load moderately on the autonomy or control factor, but have an equally large or larger loading on the method factor. In an exploratory factor analysis they formed their own dimension, illustrating that both items are referring to a related but different latent concept, strongly related with the limits imposed on activity by age and health. As we strive for an uncontaminated measure, the choice was made to exclude both items. Below are the test results for each step and the different versions of the new scale. In the best fitting solution with three factors each dimension is measured by 5 items. As only two negatively phrased items remain, we allow an error correlation between both items.

Table 5: Fit statistics for CASP15 reduced scale in Elsa wave 1

	RMSEA	CFI	TLI
One factor	.097	.928	.916
With error corr	.087	.942	.932
Two factor	.089	.940	.929
With error corr	.078	.954	.945
Three factors	.080	.952	.943
With error corr	.072	.961	.953
Second order factor	.080	.952	.943
With error corr	.072	.961	.953

This version of the factor scale has good fit characteristics and is also theoretically more delineated, as it is not contaminated with objective life circumstances such as age and health which may be important in explaining variations in quality of life.

If we exclude the same items from the 12 item scales, we find similar results for the proposed version of the scale (Wiggins et al., 2007). When trying to run second order models, cross-loadings again cause correlations higher than 1 between first and second order factors. The SHARE version, which has one item less, seems to fit better to the dual factor model. The second order model again runs into problems with the covariance matrix, due to the low discriminatory power of the dimensions control/autonomy and self-actualisation.

Table 6: Fit statistics for CASP10 and CASP9 reduced scale in Elsa wave 1

	Normal (10 items)			Share (9 items)		
	RMSEA	CFI	TLI	RMSEA	CFI	TLI
One factor	.115	.937	.919	.108	.951	.935
With error correlations	.096	.958	.944	.079	.975	.966
Two factors	.103	.952	.936	.090	.967	.955
With error correlations	.080	.971	.961	.049	.991	.987
Three factors	.093	.963	.948	.082	.975	.963
With error correlations	.077	.975	.964	.052	.991	.985
Second order factor*	.114	.942	.941	.118	.946	.922
With error correlations	.104	.954	.935	.104	.960	.939

*Problems with factor 2

In general terms the advice is to use a method factor in calculating CASP scores, and additionally to take into account the categorical nature of the data. As such not a simple sum of the items, but a categorical factor analysis should be used to wield out measurement distortion and get closer to the true score on the latent trait, quality of life. Furthermore, it is advisable to leave out 4 less useful items of the scale (a, h, i and f), if a robust, uncontaminated and subjective measure of well-being is desired, without explicit reference to objective circumstances. Depending on the available items of the scale, either a two or three dimensional form is advised.

1.1.2 CES-D

The original CES-D scale (Radloff, 1977) comprises 20 items, but shorter versions are frequently used and have shown not to lose a lot of information (Kohout, Berkman, Evans, & Cornoni-Huntley, 1993). In the European Social Survey (ESS), HRS and ELSA an 8 item version is used, while SHARE uses an 11 item version. When looking at the CES-D scale in its extended form with EFA, four sub-dimensions surface: positive affect, depressed affect, somatic complaints and interpersonal problems (Kohout et al., 1993; Radloff, 1977; Ross & Mirowsky, 1984). Looking at the 8 item version, two subscales surface, one that measures mood and one that looks at somatic aspects of depression (Van de Velde et al., 2010; Wallace et al., 2000). The scale can equally well be seen as a single scale for most applications, since internal consistency is high, and correlations between the different sub dimensions are usually higher than .90 so that they are difficult to distinguish from each other. In our case it might be relevant to look at the two factor solution, as theoretically depression, and especially it's somatic component could be linked to later life depression. Testing the scale in a CFA framework, it has also been established that the CES-D scale represents a continuum rather than forming separate factors for positively and negatively worded items, if correlations between negatively worded items are allowed (Wood et al., 2010).

Table 7: CFA of 8 item CES-D scale in ELSA wave 1

	RMSEA	CFI	TLI
1 factor	.075	.964	.958
With error correlations	.066	.978	.958
2 factors	.054	.986	.978
With error correlations	.035	.994	.991

All fit statistics, both for one or two factors, are acceptable. It is clear that allowing error correlations between negatively worded items significantly improves the models. The two factor model has a better fit, and there is a correlation of .82 between both factors. This is still a high correlation, but substantially lower than the .90 reported in the ESS. The data clearly favour the most complex model, with two factors and error correlations between the negatively worded items, but even a simple one factor model has an acceptable fit, illustrating the robust reliability of the shortened CES-D scale.

1.1.3 General health Questionnaire (GHQ)

The GHQ is a 12 item scale, intended as a general screening instrument for psychiatric morbidity (Goldberg & Williams, 1988). Most researchers looking at the factor structure of this scale have focused on the number of sub dimensions. While a large part of the scientific work has been highlighting the plausibility of a three factor structure (anxiety, social dysfunction and loss of confidence) instead of the original one factor (Graetz, 1991; Shevlin & Adamson, 2005), recently the inclusion of method effects of negative wording has shown this multidimensionality to be a measurement artefact (Hankins, 2008). The three dimensional structure groups positive and negative items in separate dimensions, and can be seen as an extension of the two factor model.

Table 8: CFA of 12 item GHQ scale in ELSA wave 1

	RMSEA	CFI	TLI
1 factor	.143	.913	.894
With error correlations	.072	.984	.973
With method factor	.087	.972	.961
2 factors (positive and neg items)	.085	.970	.963
3 factors	.076	.977	.970

Our test on a representative sample of community dwelling people aged 50 or older seems to confirm these findings, although it has to be mentioned that when a method factor is used instead of error correlations the three factor model fits better. Since the difference is quite large between the specifications with error correlations and the method factor, allowing the error correlations might be masking substantial aspects of the scale, so that it makes sense to choose the 3 factor model.

1.1.4 Satisfaction with life scale (SWLS)

The satisfaction with life scale (Diener et al., 1985) is one of the most used instruments to measure global life satisfaction. It is most commonly seen as one dimensional, comprising 5 items (Pavot & Diener, 1993), or often even just one item (Morrison, Tay, & Diener, 2011). Some researchers have found a two factor structure for the SWLS (McDonald, 1999; Wu & Yao, 2006). The last two items of the scale, which refer more to past experiences, have a different importance for the total score both in later life compared to younger people, and in different cultures compared to the US, where the scale has been most extensively tested (Hultell & Petter Gustavsson, 2008; Oishi, 2006; Pons, Atienza, Balaguer, & García-Merita, 2000). These two factors are very closely related in most studies (correlation around .90), so that a hierarchical second order factor structure is proposed. Since all items are worded in the same sense, a method factor is not necessary for SWLS. Since the scale was only included from ELSA wave 2 onwards, we tested the models on wave 2.

Table 9: CFA of 5 item SWLS scale in ELSA wave 2

	RMSEA	CFI	TLI
1 factor	0.135	0.995	0.989
2 factor	0.121	0.997	0.991
With error correlation	0.073	0.999	0.997

Both the one factor and two factor solution do not seem to fit very well according to the RMSEA, but have a very good fit according to the CFI. The two factor model fits marginally better, but the correlation between both factors is very high (.938). It is a quite surprising finding that one of the most used scales to measure subjective well-being does not fit particularly well for older respondents in England. Modification indices indicate that being satisfied with life (item c) is more closely related to evaluating one's life as ideal, and less to perceiving one's life conditions as ideal. We allow this correlation, and keep in mind that conditions seem to be less important for life satisfaction among the elderly in the UK.

1.2 Well-being measures combined: A second order measure of well-being

Now that we have an indication on the structure of separate aspects of well-being, it can be investigated to what extent these different aspects coincide. From the theory a number of specific hypotheses on the structure of wellbeing can be deducted. The most influential approach in wellbeing research, hedonic well-being, based on the work of Diener, assumes two components, an affective part, captured in our measurements by the CES-D and GHQ scales and a cognitive part, in our available data the SWLS (Diener, 1984; Diener et al., 1999; Pavot & Diener, 1993). An alternative approach states that wellbeing results not from happy mood or a good evaluation of one's life, but through attainment of life goals. This eudaimonic approach does not directly address it's relation with hedonic measures, but sees itself as a separate and conceptually different approach to measuring wellbeing. This raises the question if and to what extent both approaches are different from one another. Using the best fitting structural form for each sub-dimension can help us investigate the dimensionality of well-being, by looking at second order structures.

Table 10: Overview of scales present in each wave of ELSA

	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Hedonic Affective					
CES-D	X	X	X	X	X
GHQ12	X		X		X
Hedonic Cognitive					
SWLS		X	X	X	X
Eudaimonic					
CASP	X	X	X	X	X

Two aspects of our available data complicate this undertaking. Firstly, not all scales are available in each wave of ELSA (Table 10). Therefore we will present the findings of the analysis on the most complete set of measures, in wave 3, in the text.[5] Secondly, each scale allows a different way of answering an item. The CES-D items are binary in nature, while both CASP and GHQ provide 5 response categories and the SWLS 7. Although the CASP and GHQ items have the same number of answering possibilities, their meaning differs as in CASP the frequency of something happening is asked for, while in GHQ the respondent is asked to compare the something with their 'usual' behaviour. This means that the highest correlations between sub-dimensions will logically occur within the same scale. It is nonetheless useful to look at the correlations between the sub-dimensions to investigate which aspects of well-being are less related. Correlations lower than .60 are considered weak, while correlations higher than .75 are considered strong.

[5] Preliminary analysis on waves 1 and 2 point to similar results.

Table 11: Overview of correlations between subdimensions in wave 3 of ELSA (n=8598)

	SWLS present	SWLS past	GHQ anxiety	GHQ social dysfunction	GHQ loss of confidence	CES-D somatic	CES-D mood	CASP Control Autonomy	CASP Self-Realisation
SWLS past	0.926								
GHQ anxiety	0.609	0.527							
GHQ social dysfunction	0.579	0.454	0.715						
GHQ loss of confidence	0.593	0.521	0.857	0.748					
CES-D somatic	-0.508	-0.384	-0.664	-0.630	-0.617				
CES-D mood	-0.628	-0.544	-0.766	-0.703	-0.674	0.834			
CASP Control Autonomy	-0.730	-0.641	-0.666	-0.589	-0.717	0.640	0.656		
CASP Self-Realisation	0.779	0.753	0.546	0.617	0.607	-0.667	-0.639	-0.847	
CASP Pleasure	0.710	0.732	0.582	0.566	0.648	-0.512	-0.646	-0.790	0.886

The 10 factor model, specifying all subscales of all available measures of well-being, and error correlations between negatively worded items within each scale, has a good fit (Table 12). The reduced 15 item version of CASP was used in this analysis. Before looking at the second order structure of well-being, it is relevant to examine the correlations in detail. As expected, the highest correlations can be observed between subscales derived from a similar instrument. More relevant for the topic of this paper, is that a number of concepts only are weakly related to each other. Satisfaction with life in general can be seen as only weakly related to most aspects of mental health, which is indicated by the moderate correlations with most subscales of the GHQ and CES-D. On the other hand satisfaction with life, especially in the present, is strongly related to self-actualisation. Anxiety is closely related to symptoms of a depressive mood, but less to self-actualisation and pleasure. Loss of confidence seems closely associated with low control and autonomy. Somatic symptoms of depression are especially weakly related to satisfaction with past life, and only moderately with satisfaction with life in the present, or pleasure. In general depressed mood is slightly closer related to satisfaction with life and general mental health compared to somatic symptoms. Surprisingly the pleasure domain of CASP is not more strongly related to the hedonic measures in comparison with the domains control and autonomy and self-realisation. This could indicate that for most respondents, enjoyment is something else than mere satisfaction. Similarly, no measure of positive affect is available in the current ELSA dataset, so it could well be that pleasure is not that closely related to negative affect and satisfaction with life, but more with positive affect.[6] A second explanation is that the frequency of enjoyment asked for in CASP is more related to eudaimonic aspects of well-being than satisfaction with current or past life. A last explanation is that eudemonia, or fulfilling one's psychological needs, is enjoyable and as such should always be seen as partly hedonic.

[6] Positive affect apparently is available for a subsample of ELSA respondents, who provided saliva samples, through ecological momentary assessments derived from their logbooks (Steptoe & Wardle, 2011). These data are not part of the current version of the ELSA dataset.

Table12: Second order CFA in wave 3 of ELSA

	RMSEA	CFI	TLI
10 single order factors	0.052	0.954	0.948
1 second order factor	0.074	0.903	0.896
2 second order factor	0.070	0.914	0.907
3 second order factor[7]	0.053	0.950	0.946
4 second order factor	0.053	0.951	0.947

In a second step we will investigate what second order factor structure fits best. In theoretical terms this can be seen as an empirical test of the nature of subjective well-being. In the single order model, all sub-dimensions are allowed to correlate with one another. Specifying a second order factor means reducing all these aspects of well-being to a single dimension. Although this model has an acceptable fit in terms of RMSEA, this seems less the case for the other fit indices. This means that a single well-being concept is defendable, but does not fully grasp the complexity of the subject at hand. Two second order factors, hedonic and eudaimonic wellbeing, do not greatly improve our model. This means that the division between hedonic and eudaimonic measures is not that substantial. It has to be kept in mind that pleasure is seen as a eudaimonic measure in this context.[8] Specifying a dimension of cognitive, affective and eudaimonic wellbeing on the other hand, fits our data remarkably well. An extra factor for the two measurements of affective wellbeing does not significantly improve our model, so that we can confidently assume a three dimensional nature of wellbeing.

In very general terms it can be said that satisfaction with life is not that closely related to affective elements of hedonic well-being, but is quite closely associated with eudaimonic well-being in general and self-realisation in specific. Eudaimonic well-being in itself is both strongly related to affective and cognitive aspects of well-being.

[7] One cross loading had to be allowed in this model to avoid a negative covariance of self-realisation with cognitive hedonic well-being. The item "I feel satisfied with the way my life has turned out" of the self-realisation domain was allowed to load on the hedonic cognitive latent second order factor. This is defendable since the nature of the item explicitly refers to satisfaction with life.

[8] An alternative model with pleasure as a part of hedonic wellbeing did not converge, indicating a worse model specification.

2.2 Measurement equivalence over subgroups

In a second step the equivalence of our measurement over subpopulations will be investigated. Looking at measurement equivalence questions the often implicit assumption that latent constructs are measured in the same way across groups or countries. We investigate if constructs can be compared in a meaningful way across these groups, so that differences between group scores can be attributed to differences in the latent concept, and not to measurement issues. Several sources of measurement interference can be distinguished. A given measure could be interpreted in a conceptually different way by different ethnic, social or national groups. In a cross-national framework, the fact that a measurement instrument is translated in different languages could cause different interpretations of the latent concept. Measurement issues can also indicate substantial differences in how different groups within a country relate to a concept. It could be for example, that men and women interpret an item in a different way, or that differences in educational level have an effect on measurement. Again it is therefore important to remember that a failure to establish measurement equivalence does not mean that a scale is useless, or that previous analysis using a scale is invalid. It should urge researchers to approach differences between subgroups with care, and to highlight different ways in which the latent concept is understood by different groups.

2.2.1 Method

In practice, measurement invariance can be tested using two different techniques, CFA, which we already used in the first part of the analysis, and item response theory (IRT) (Raju, Laffitte, & Byrne, 2002; Reise, Widaman, & Pugh, 1993). The most important difference between both methods in substantial terms, is that CFA assumes a linear relationship between an item and the underlying construct, while in IRT a non-linear relationship is assumed. Both methods lead to similar substantive results, and examine measurement invariance as the invariance of the relationship between latent construct and true item score across subpopulations (Raju et al., 2002; Reise et al., 1993). In this study the CFA approach will be used to test measurement equivalence, as it is more commonly used to investigate invariance of polytomous items and multidimensional latent concepts.

Multiple group confirmatory factor analysis is a rigorous technique for such an analysis (Brown, 2006; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000) . Measurement equivalence consists of different levels, each of which can be seen as a cumulative step of comparability, associated with constraining a set of parameters in the CFA model. Each level has a meaning in terms of how comparable a scale is among subgroups. Disturbances to the levels of measurement equivalence can either be due to substantial issues such as a different meaning of a concept, or to measurement issues such as differences response styles across groups. Inappropriate sampling procedures, translation errors or coding blunders may equally be responsible for non-invariance, but are very hard to detect.

Dimensional invariance exists if the same number of dimensions surface from a measurement instrument across groups. In step 1 we have already illustrated that the number of dimensions on a complete sample can already pose a number of complications, when some items are worded negatively. Similarly, when a scale comprises several closely related factors, they can be more closely related in some countries than in others, so that a choice has to be made that fits all countries.

Configurational invariance exists if non salient factor loadings are equal to zero in all groups. This can be seen as a basic model, which checks if the same items load on the same factors in subgroups. In practice most tests of invariance start by comparing groups at this step, as dimensional invariance is usually assumed. Configurational invariance is examined by examining if the theoretical model fits the data in each country separately to more or less the same extent.

Metric or pattern invariance exists if these salient factor loadings are all equal among subgroups. Each item then can be seen as having the same contribution to the latent concept in all subgroups. One possible reason for the absence of metric invariance is the presence of extreme response styles in one of the subgroups (Baumgartner & Steenkamp, 2001; G. W. Cheung & Rensvold, 2000), the other possibility is that the latent concept has a different meaning to the group under study (Gregorich, 2006). When metric invariance is established, factor variances and covariances can be compared between groups.

Scalar or strong invariance exists if the intercepts or thresholds of items are equal in subgroups. Differential additive response styles are seen as the main explanation in terms of measurement bias for the lack of this level of equivalence (Baumgartner & Steenkamp, 2001; G. W. Cheung & Rensvold, 2000). One example of differential additive response is that in different cultures the same item response might mean something else due to social desirability. Scalar invariance means that observed and factor means can be compared between groups.

These levels of invariance do not have to be satisfied absolutely on all items. Partial invariance can also be assessed, by freeing the relevant parameter for a separate item (Byrne, Shavelson, & Muthén, 1989; Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). When partial invariance is established, only the invariant items should be used to compare subgroups on the latent dimension.

Depending on the estimation method used, a number of test statistics are available to see if the level of equivalence is supported by the data. As maximum likelihood procedures allow a better model evaluation and comparison, equivalence is investigated defining the items as interval while using ML estimation. To check if the model is also valid when taking account of the ordinal nature of the data, the final models are tested in a WLSMV framework and the model fit is evaluated again (Davidov, 2008).

In a final step, the latent means of the reduced CASP scale will be compared with the observed ones by country, as a robust check of measurement invariance. The latent means control for different item functioning or different meanings in different countries, while the observed scores do no. If only small differences in country ranking surface, it illustrates the invariance of measurement, while bigger differences in ranking between the observed and latent scores, illustrate measurement variance.

2.2.2 Measurement invariance across countries

Measurement invariance can be assessed across a number of subpopulations. A first and obvious check for the validity of comparisons is assessing equivalence across gender, age groups and educational level within a cross-sectional survey. Are differences in wellbeing between different age groups due to differences in answering the questionnaire, or are they genuine and substantial

differences? A second important step is looking at longitudinal equivalence of a scale over different waves of a panel study. This kind of analysis investigates if people get used to a questionnaire and change their answering behaviour, or if the change over time if a change in the true value of the latent concept. A third and the most well-known possibility for measurement equivalence is assessing the structure of a latent concept over several countries.

Since it would lead us too far to test for all of these forms of equivalence for all of the scales, we have to make a selection. Since it is a relatively specific measure, present in a number of comparable international studies, on which not much investigation of equivalence has been done, we will focus on the CASP scale in SHARE, a database meant to compare between countries. In the previous step we have identified a reduced CASP scale that satisfies both strict theoretical and methodological criteria, and as such we will investigate this scale. To investigate cross-country comparability using SHARE, only 12 items are available, of which only 9 remain in our reduced measure.

In the first part of the analysis, it was illustrated with wave 1 of the ELSA dataset that a two factor model fits better when using only 9 items (Table 5). This is also the case when using wave 2 of the SHARE dataset.[9] As such we will test if a bi factor model for CASP, accounting for negative wording in two items, is invariant across Europe. We use wave 2 as a larger amount of countries are included wave 1 and 3, and as such there is a greater variability of countries. A first step is to test the two factor model separately in each country (Table 13). The data fit reasonably well in all countries, only Austria and Poland show a moderate fit rather than a good one. In substantial terms this means we can assume the 9 items are captured in two dimensions across Europe.

Table 13: Model fit of two factor model in each country of wave 2 of SHARE

	RMSEA	CFI	SRMR	Chi sq
Austria	.065	.966	.032	164.544
Germany	.055	.968	.027	212.905
Sweden	.040	.982	.019	129.517
Netherlands	.041	.981	.022	133.706
Spain	.056	.974	.028	191.892
Italy	.059	.970	.040	282.280
France	.043	.977	.025	156.179
Denmark	.034	.989	.019	100.568
Greece	.051	.978	.024	232.619
Switzerland	.037	.981	.023	73.679
Belgium	.051	.973	.027	225.384
Czech republic	.045	.983	.024	167.485
Poland	.071	.968	.036	329.403
Ireland	.056	.965	.033	111.686

In table 14 the cumulative models of invariance are applied to the data. The first model examines if the items are connected to the same latent concepts, and if we can consider the model configurally equivalent. In contrast to the preliminary test of dimensional invariance (Table 12), this is done by

[9] 2 factor model in share: RMSEA=.039, CFI=.985, SRMR =.018
3 factor model in share: RMSEA=.050, CFI=.975, SRMR =.027

including all the countries in one model, with factor loadings, intercepts and factor means allowed to vary for every country. A marker variable for every latent concept, needed to identify the model, is constrained to one across all countries. It is clear that a model distinguishing eudaimonic from hedonic well-being loads on the same items across all countries, since the fit is good (RMSEA<.06, CFI>.95, SRMR<.05).

Table 14: Results of ML CFA of two factor model of CASP and testing for measurement invariance in wave 2 of SHARE

	RMSEA	CFI	SRMR	Chi sq	df
Configural invariance	.051	.976	.027	2511.847	350
Metric	.064	.952	.073	4772.290	441
Partial metric	.056	.963	.053	3758.611	440
Scalar	.099	.859	.113	13149.847	531
Partial scalar	.059	.953	.055	4703.630	497

If the two concepts of well-being are understood similarly in the different countries, each factor loading is equal to that of the same item in the other countries. Full metric invariance is only moderately supported by the data. Freeing up the factor loading of the item "I look forward to every day" in Italy improves the model significantly, and as such establishes partial measurement equivalence. When it is allowed to vary, the loading of this item in Italy changes sign and diminishes significantly in size.[10] This means that for Italian respondents, the frequency of looking forward to a new day is not related (or slightly negative) to the concept of pleasure and enjoyment as defined in CASP, and is very different from the other European countries. This might relate to a nuance of translation into Italian, or how Italians interpret the item. The statement that one often looks forward to a new day could be interpreted as not enjoying today, which would explain the negative loading on the pleasure factor. Another explanation could be that Italian respondents answered this item in a more extreme way than respondents from other countries. The last and most straightforward interpretation is that how often one looks forward to the next day does not matter is not related to how much an Italian enjoys his life. The fact that we have full metric equivalence for all other countries means that we can safely examine the constructs correlates with each other or other variables of interest. When Italy is included in the comparison, this item has to be left out of the construct, or allowed to vary by country.

A final form of invariance, which allows for comparison of latent means, is scalar invariance. This means that not only the loadings, but also the relation of the intercept (or item mean) with the latent construct is equal across countries. In practice full scalar invariance is more the exception than the rule, and also in our model there is no full scalar invariance. By allowing 33 intercepts of the total of 126 to vary, the scalar equivalence model achieves a good model fit. While for Switzerland, no intercepts had to be freed, for Greece 5 out of 9 had to be adjusted. The intercepts that had to be freed most often are for the items "feeling left out of things" and "doing the things you want to do". In countries located more in the south of Europe, such as Italy, France and Greece people on average had the feeling being left out more, while in Germany and the Netherlands people had this feeling less frequent. Similarly, in Greece, Italy and Spain people feel that they can do less often the things they want to, while in Scandinavian countries people on average do things they want to more often.

[10] The standardized factor loading for this item is -.187 in Italy, compared to .921 in the all other countries

This suggests their might be broad cultural norms influencing response behaviour, in the sense that in Northern European countries people tend to report feeling left out slightly less frequent, and reporting more frequent that they are able to do what they want, while Southern Europeans tend to report more frequent feelings of being left out and less frequent doing things they want for a similar score on the latent CASP factors.

To test if our findings also hold when we consider the items as ordinal, the analysis was replicated using the same model specifications, but defining the items as ordinal, and using WLSMV estimation (Table 15). Instead of a single intercept for each item, in this specification thresholds on a latent continuous scale for each item are used to discriminate between answering categories. As such an item with four categories is defined by one loading (and an associated scale factor) and three thresholds. Because the ordinal model is considerably stricter, we attach more importance to the CFI, and are satisfied with an acceptable model fit for RMSEA (<.08). Again configural and partial metric equivalence can be assumed for the whole set of countries. To achieve partial scalar equivalence, the loading of item j had to be released in Belgium, which was substantially lower than in other countries.[11]

Table 15: Results of WLSMV CFA of two factor model of CASP and testing for measurement invariance in wave 2 of SHARE

	RMSEA	CFI	TLI	df
Configural	.058	.987	.981	350
Metric	.108	.942	.934	441
Partial metric	.074	.973	.969	440
Scalar	.107	.901	.935	765
Partial scalar	.081	.950	.962	669
+ freeing loading j in Belgium	.076	.957	.967	667

Now we know that CASP, seen as a two-dimensional measure for subjective well-being, is partially equivalent across Europe. We can compare the latent means of both factors in different European countries assuming that these differences reflect real differences, and are not the result of measurement bias. As an illustration, we plotted the differences in latent means on both factors comparing the average level of wellbeing in other European countries with Germany (Figure 2).

[11] Because of the multiple thresholds, metric and scalar equivalence tests are closely associated when using categorical items (Muthén & Muthén, 2010, 433). The different levels of equivalence as such partly loose their definition in the sense that achieving scalar equivalence can mean freeing item loadings instead of thresholds, as in this case.

Figure 2: Comparison of latent means in European countries using the two dimensional CASP standardised factor score

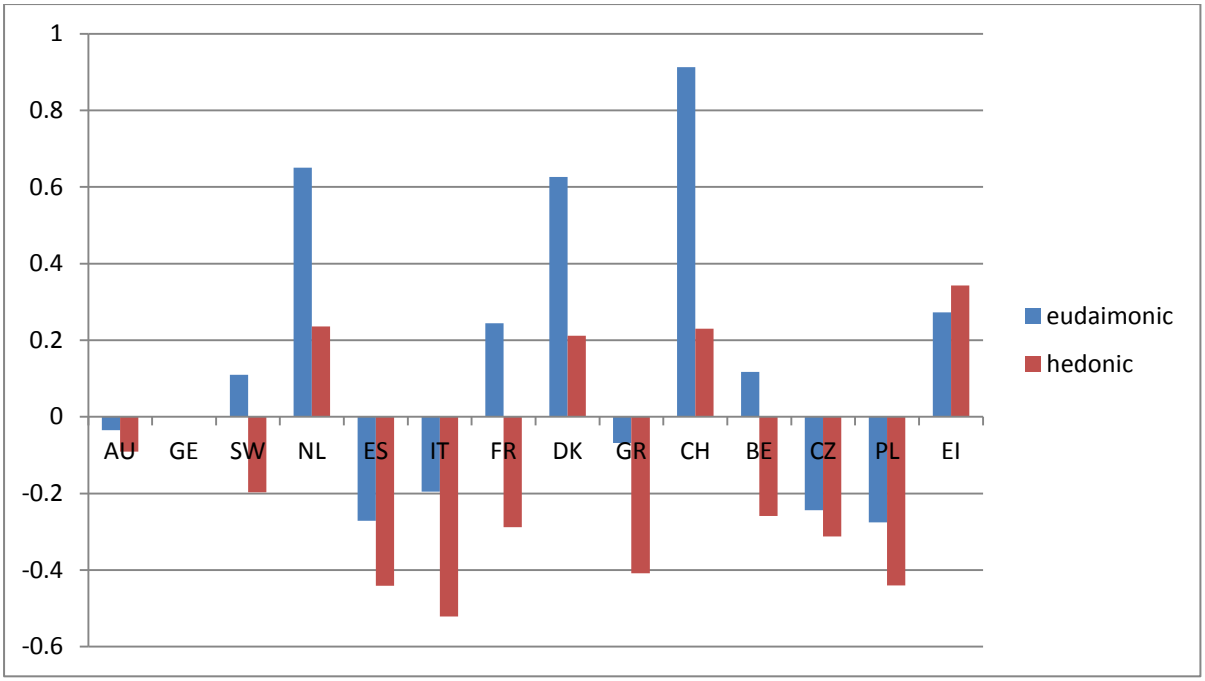


Table 16: Difference in standardized factor scores between Germany and other countries (MLE CFA)

	Eudaimonic Factor		Hedonic Factor	
	Score Difference	S.E.	Score Difference	S.E.
Austria	-0.035	0.038	-0.091	0.013
Germany	0		0	
Sweden	0.11	0.039	-0.197	0.038
Netherlands	0.65	0.035	0.236	0.035
Spain	-0.271	0.033	-0.441	0.032
Italy	-0.195	0.029	-0.521	0.03
France	0.244	0.033	-0.288	0.032
Denmark	0.626	0.035	0.212	0.035
Greece	-0.068	0.033	-0.408	0.031
Switzerland	0.913	0.048	0.23	0.044
Belgium	0.117	0.031	-0.259	0.028
Czech Republic	-0.244	0.033	-0.312	0.029
Poland	-0.275	0.03	-0.44	0.028
Ireland	0.273	0.045	0.343	0.045

A first observation that can be made is that the differences in eudaimonic well-being are larger the differences in hedonic well-being. This is explained by the fact that the hedonic subscale only had three items, and as such has a smaller variation. In most countries hedonic and eudaimonic wellbeing consistently deviate in the same direction. Sweden, France and Belgium are exceptions to this pattern. In general terms it can be said that countries in the South or East of Europe have lower

levels of both hedonic and eudaimonic wellbeing than Germany. Wellbeing is markedly higher in the Netherlands, Denmark, Switzerland and Ireland in later life.

Comparison of observed and latent means

Our analysis suggests partial scalar measurement invariance of CASP, seen as a two dimensional measure encompassing both hedonic and eudemonic aspects of subjective well-being, across Europe. As a test of robustness we will compare the ranking of countries on two versions of this scale, on the one hand the simple observed sum score (of the 15 items), which is the way CASP is proposed to be used (Hyde et al., 2003) and on the other hand the latent means of the partial scalar CFA model (with WLSMV estimation), which was the final step in our analysis. These latent means can be regarded as free of measurement bias induced by country context or question wording. Comparing how countries score on both scales as such is a robust test of the invariance of the scale. Since the units of the observed and latent scale can not be compared in a meaningful way, the ranking of countries according to their mean score is examined for each sub dimension of CASP (Table 17 and Table 18).

Table 17: Comparison of observed and latent means of Eudaimonic factor in CASP

	Observed means		Latent means		Ranking
Country	Mean	S.D.	Mean	S.D.	
Austria	12.409	3.723	-0.027	0.466	9->9
Germany	12.955	3.525	-0.022	0.439	6->8
Sweden	13.312	3.071	0.025	0.430	4->7
Netherlands	14.137	3.178	0.306	0.446	2->2
Spain	11.599	3.768	-0.139	0.454	10->13
Italy	11.077	4.083	-0.123	0.488	14->11
France	12.832	3.673	0.115	0.458	7->4
Denmark	14.051	3.167	0.283	0.457	3->3
Greece	11.138	3.603	-0.039	0.482	12->10
Switzerland	14.412	2.908	0.375	0.412	1->1
Belgium	12.653	3.707	0.036	0.464	8->6
Czech Republic	11.122	3.539	-0.133	0.413	13->12
Poland	11.368	4.148	-0.172	0.516	11->14
Ireland	13.183	3.321	0.094	0.432	5->5

Note: Spearman rank order correlation = .87 with p=.0016

Table 18: Comparison of observed and latent means of Hedonic factor in CASP

	Observed means		Latent means		Ranking
Country	Mean	S.D.	Mean	S.D.	
Austria	7.696	1.602	1.419	0.760	7->6
Germany	7.815	1.443	1.520	0.688	5->5
Sweden	7.761	1.454	1.362	0.637	6->7
Netherlands	7.928	1.553	1.709	0.651	4->3
Spain	6.963	1.886	1.046	0.730	11->12
Italy	5.800	1.708	0.998	0.779	14->14
France	6.586	1.985	1.259	0.636	13->8
Denmark	8.262	1.296	1.822	0.742	1->2
Greece	7.069	1.658	1.164	0.667	10->11
Switzerland	8.070	1.370	1.692	0.652	3->4
Belgium	6.933	2.150	1.209	0.819	12->10
Czech Republic	7.259	1.657	1.211	0.724	8->9
Poland	7.074	1.939	1.043	0.795	9->13
Ireland	8.245	1.326	1.840	0.752	2->1

Note: Spearman Rank order correlation .88 with p=.0014

For both subscales of CASP, observed and latent means point towards similar differences between countries. Although the ranking was not exactly the same, both operationalizations of the CASP sub domains capture the differences in mean well-being between countries in a very similar way, as is shown by the high Spearman rank order correlations, respectively .87 and .88. The countries for whom the rank differed most for the eudaimonic factor where France, Poland, Sweden, Spain and Italy, while for the hedonic factor again France and Poland had the largest difference in rank.

3. Conclusions

This paper investigates the empirical measurement of well-being in later life, by examining a number of commonly used scales and looking at their interrelations. This examination is framed in the discussion on the difference between hedonic and eudaimonic well-being. The dominant approach, hedonic well-being, assumes that well-being emanates from pleasure and the avoidance of painful experiences, however these are defined by the individual. Measuring wellbeing in this framework tries to capture moods and emotions on one hand, in the form of positive and negative affect, and cognitive evaluations of one's life on the other hand (Diener, 1984). Eudaimonic well-being is not such a unified approach as hedonic well-being, and consists of several multidimensional approaches (Hyde et al., 2003; Ryan & Deci, 2000; Ryff & Keyes, 1995). What they have in common is that they assume well-being emerges as a result of the satisfaction of universal human psychological needs. While Ryan & Deci (2001) and Hyde et al. (2003) assume pleasure, or hedonic well-being, is one of those needs, Ryff & Keyes (1998) state that at best there is a weak relation between need fulfilment and pleasure.

To what extent do indicators of these different aspects of well-being, commonly developed by testing on either relatively small groups of students or in population wide large scale surveys, replicate their structure among adults aged 50 or older in England? Both instruments aimed at capturing negative affect, CES-D and GHQ, performed most in line with their expectations. While considering CES-D as a one dimensional instrument screening for depression is acceptable, a more fine grained approach to depression clearly distinguishes somatic aspects from emotional ones. The GHQ measure in a similar vein is acceptable as a one dimensional construct, but allows more nuance when looking at anxiety, social and confidence aspects of psychological morbidity separately. Satisfaction with life, the most commonly used measure for well-being, seems to perform relatively poorly. Not only can a distinction between satisfaction with the past or present be made, which was already noted by other researchers (Hultell & Petter Gustavsson, 2008; Oishi, 2006), in our sample satisfaction or seeing one's life as ideal was less related with how one perceives his life conditions.

The most challenging scale was CASP, which was developed specifically for adults aged 50 and over and originally tested using wave 1 of ELSA. A reliable and robust measurement of subjective quality of life, as intended by the developers, is possible with this scale, if it is used in an adapted and shortened form. The main problems of CASP in its original version were a number of weakly loading items, of which one was still present in the advised 12 item version, next to the presence of concepts, such as autonomy, control, and self-realisation, which are too closely related to be seen as independent. Two of the superfluous items related to the limitations imposed by age and health, and seemed to define a separate dimension, less strongly related to wellbeing, bringing to mind the concept of frailty. The theoretical foundation of the scale relies on the view that "any QOL measure should be distinct from contextual and individual phenomena that might influence it, such as health, social networks and material circumstance" (Hyde et al., 2003, 187). Therefore it is somewhat inconsistent that the items measuring the influence of exactly these limitations were present in both the original instrument (age, health, family responsibilities and money) and the revised one (age and money). Since all subsequent steps of analysis rely on a theoretically robust and methodologically sound scale, a new version of CASP comprising either 15 items (derived from CASP19), 10 items (derived from CASP12) or 9 items (derived from CASP12 in SHARE) was developed. In both the 15

and 10 item versions, three sub dimensions, control and autonomy, self-realisation and pleasure, surface, while in the limited 9 item SHARE version only two dimensions surfaced. These two dimensions reflected the split between hedonic and eudaimonic aspects of well-being.

The relations between these different facets of well-being were largely in line with our expectations. Present satisfaction with life was slightly closer related to measures of negative affect, control and autonomy and self-realisation than satisfaction with the past life. Both present and past satisfaction were more related to aspects of human flourishing than to psychological morbidity and depression. Anxiety, social dysfunction, pleasure and both dimensions of satisfaction were more related to emotional symptoms of depression than somatic ones, while the associations were about the same for control and self-realisation. Surprisingly pleasure was not significantly closer related to both affective and evaluative aspects of hedonic well-being compared with other dimensions of the CASP scale. Looking at the second order structure of the scales, it is clear that the difference between hedonic and eudaimonic well-being had been exaggerated in the literature. If a multidimensional concept of wellbeing is used, it seems clear that a threefold structure, distinguishing cognitive, affective and eudaimonic well-being is more informative.

Can eudaimonic well-being in later life be measured across Europe in a reliable way? Our analysis, departing from a dual factor model of the CASP scale suggests this is at least partially the case. Conceptually well-being is measured by the same items in all countries, except Italy and Belgium, where looking forward to the next day is less related with control, autonomy and self-realisation than in other countries. Next to this partial metric equivalence, partial scalar equivalence could also be established. The deviations in answering patterns found in the intercepts and thresholds of the items suggest that different cultural sensitivities exist in the North and South of Europe regarding social inclusion and individual decisions in later life. In the South feelings of being left out were reported more, and people felt they were doing less what they wanted, than in the North. But although some items were sensitive to these differences, when taking the whole scale into account it can be safely assumed that the latent means reflect real differences and not just measurement artefacts.

What would help us answer the questions posed in this analysis better, or in other words what are the suggestions for further research? First of all, access and inclusion to more measures of well-being, such as positive affect and perhaps loneliness could broaden our understanding of how eudaimonic well-being relates to cognitive and affective aspects. In the case of loneliness this creates the question to which extent it should be seen as an aspect of well-being, and hence a basic psychological need, instead of a possible cause of low well-being, and hence a driver.

Bibliography

Alexopoulos, G. S. (2005). Depression in the elderly. *Lancet*, *365*(9475), 1961–70. doi:10.1016/S0140-6736(05)66665-2

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research*, *38*(2), 143–156. Retrieved from http://www.jstor.org/stable/10.2307/1558620

Beauducel, A., & Herzberg, P. Y. (2006). Structural Equation Modeling : A On the Performance of Maximum Likelihood Versus Means and Variance Adjusted Weighted Least Squares Estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203.

Beaumont, J. (2011). *Measuring National Well-being - Discussion paper on domains and measures*. London.

Beekman, a T., & Deeg, D. J. (1995). Major and minor depression in later life: a study of prevalence and risk factors. *Journal of Affective Di*, *36*, 65–75. doi:10.1111/j.1752-0606.2011.00243.x

Booth, P. (2012). *… and the Pursuit of Happiness. Wellbeing and the Role of Government*. (P. Booth, Ed.). London: Institute of Economic Affairs.

Bradburn, N. M. (1969). *The structure of Psychological Well-being*. Chicago: Aldine .

Brickman, P., & Campbell, D. T. (1971). Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation level theory A symposium* (pp. 287–302). Academic Press.

Brickman, P., Coates, D., & Janoff-Bulman, R. (1978). Lottery winners and accident victims: is happiness relative? *Journal of personality and social psychology*, *36*(8), 917–27. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/690806

Brown, T. A. (2003). Confirmatory factor analysis of the Penn State Worry Questionnaire: Multiple factors or method effects? *Behaviour Research and Therapy*, *41*(12), 1411–1426. doi:10.1016/S0005-7967(03)00059-7

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.

Bury, M. (1995). Aging, gender and sociological theory. In S. Arber & J. Ginn (Eds.), *Connecting Gender and Ageing: A sociological approach*. Philadelphia: Open University Press.

Byrne, B. M., Shavelson, R. J., & Muthén, B. O. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures : The Issue of Partial Measurement In variance, *ID*(3), 456–466.

Börsch-Supan, A., & Jürges, H. (Eds.). (2005). *The Survey of Health , Aging , and Retirement in Europe – Methodology*. Mannheim: MEA.

Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The Quality of American Life: Perceptions, Evaluations, and Satisfactions.* (p. 583). New York: Russell Sage Foundation. Retrieved from http://www.amazon.com/Quality-American-Life-Satisfactions-Publications/dp/0871541947

Charles, S. T., Reynolds, C. A., & Gatz, M. (2001). Age-related differences and change in positive and negative affect over 23 years. *Journal of Personality and Social Psychology*, *80*(1), 136. Retrieved from http://psycnet.apa.org/journals/psp/80/1/136/

Chen, Y., Rendina-gobioff, G., & Dedrick, R. F. (2010). Factorial Invariance of a Chinese Self-Esteem Scale for Third and Sixth Grade Students : Evaluating Method Effects Associated with Positively and Negatively Worded Items. *International journal of educational and psychological assessment*, *6*(December), 21–35.

Cheung, G. W., & Rensvold, R. B. (2000). Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Research Using Structural Equations Modeling. *Journal of Cross-Cultural Psychology*, *31*(2), 187–212. doi:10.1177/0022022100031002003

Cheung, Y. B. (2002). A confirmatory factor analysis of the 12-item General Health Questionnaire among older people. *International journal of geriatric psychiatry*, *17*(8), 739–44. doi:10.1002/gps.693

Crawford, J. R., & Henry, J. D. (2004). The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. *The British journal of clinical psychology / the British Psychological Society*, *43*(Pt 3), 245–65. doi:10.1348/0144665031752934

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. (H. Collins, Ed.)*Annals of Physics* (Vol. 54, p. 303). Harper & Row. Retrieved from http://www.amazon.com/Flow-Psychology-Experience-Mihaly-Csikszentmihalyi/dp/0060920432

Davidov, E. (2008). Measurement Equivalence of Nationalism and Constructive Patriotism in the ISSP: 34 Countries in a Comparative Perspective. *Political Analysis*, *17*(1), 64–82. doi:10.1093/pan/mpn014

Demakakos, P., McMunn, A., & Steptoe, A. (2010). Well-being in older age: a multidimensional perspective . In J. Banks, C. Lessof, J. Nazroo, N. Rogers, M. Stafford, & A. Steptoe (Eds.), *Financial circumstances, health and well-being of the older population in England. The 2008 English Longitudinal Study of Ageing.* (pp. 115–177). London: Institute for fiscal studies. Retrieved from http://www.ifs.org.uk/elsa/report10/ch4.pdf

DiStefano, C., & Motl, R. W. (2009). Self-Esteem and Method Effects Associated With Negatively Worded Items: Investigating Factorial Invariance by Sex. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(1), 134–146. doi:10.1080/10705510802565403

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, *95*(3), 542–575. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21228133

Diener, E., & Diener, C. (1996). Most people are happy. *Psychological science*, *7*(3), 181–185. Retrieved from http://pss.sagepub.com/content/7/3/181.short

Diener, E., Emmons, R., Larsen, R. J., & Griffin, S. (1985). Satisfaction with life scale. *Journal of personality assessment*, *49*(1), 71–75. Retrieved from http://www.unt.edu/rss/SWLS.pdf

Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: revising the adaptation theory of well-being. *The American psychologist*, *61*(4), 305–14. doi:10.1037/0003-066X.61.4.305

Diener, E., Sapyta, J. J., & Suh, E. (1998). Subjective Well-Being Is Essential to Well-Being. *Psychological Inquiry*, *9*(1), 33–37.

Diener, E., Suh, E., & Lucas, R. E. (1999). Subjective well-being: Three decades of progress. *Psychological bulletin*, *125*(2), 276–302. Retrieved from http://psycnet.apa.org/journals/bul/125/2/276/

Dolan, P., & Peasgood, T. (2008). Measuring Well-Being for Public Policy: Preferences or Experiences? *Journal of Legal Studies*, *37*(2), 5–31.

Erikson, E. (1959). Identity and the life cycle. *Psychological Issues*, *1*(1), 18–164.

Fried, L. P., Tangen, C. M., Walston, J., Newman, a B., Hirsch, C., Gottdiener, J., Seeman, T., et al. (2001). Frailty in older adults: evidence for a phenotype. *The journals of gerontology. Series A, Biological sciences and medical sciences*, *56*(3), M146–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11253156

Furedi, F. (2004). *Therpay Culture: Cultivating vulnerability in an uncertain age*. London: Routledge.

Goldberg, D. P. (1988). A User's Guide to the GHQ. Windsor. Retrieved from http://www.gl-assessment.co.uk/health_and_psychology/resources/general_health_questionnaire/faqs.asp?css=1

Goldberg, D. P., & Williams, P. (1988). *A user's guide to the General Health Questionnaire.* Basingstoke.

Graetz, B. (1991). Psychiatric Epidemiology Multidimensional properties of the General Health Questionnaire *, 132–138.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurmenet invaraicne using the confirmatory factor analysis framework. *Med Care*, *44*(11), 78–94.

Hankins, M. (2008). The reliability of the twelve-item general health questionnaire (GHQ-12) under realistic assumptions. *BMC public health*, *8*, 355. doi:10.1186/1471-2458-8-355

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A*, *6*(1), 1–55. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10705519909540118

Hultell, D., & Petter Gustavsson, J. (2008). A psychometric evaluation of the Satisfaction with Life Scale in a Swedish nationwide sample of university students. *Personality and Individual Differences*, *44*(5), 1070–1079. doi:10.1016/j.paid.2007.10.030

Hyde, M., Wiggins, R. D., Higgs, P., & Blane, D. (2003). A measure of quality of life in early old age: the theory, development and properties of a needs satisfaction model (CASP-19). *Aging & mental health*, *7*(3), 186–94. doi:10.1080/1360786031000101157

Kahneman, D., & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of economic perspectives*, *20*(1), 3–24.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: the day reconstruction method. (F. A. Huppert, B. Kaverne, & N. Baylis, Eds.)*Science*, *306*(5702), 1776–80. doi:10.1126/science.1103572

Kahneman, D., & Thaler, R. H. (2006). Utility Maximization and Experienced Utility. *Journal of economic perspectives*, *20*(1), 221–234.

Kercher, K. (1992). Assessing Subjective Well-Being in the Old-Old: The PANAS as a Measure of Orthogonal Dimensions of Positive and Negative Affect. *Research on Aging*, *14*(2), 131–168. doi:10.1177/0164027592142001

King, D. A., & Markus, H. E. (2000). Mood disorders in older adults. In S. K. Whitbourne (Ed.), *Psychopathology in later adulthood* (pp. 141–172). New York: Wiley.

Kohout, F. J., Berkman, L. F., Evans, D. a., & Cornoni-Huntley, J. (1993). Two Shorter Forms of the CES-D Depression Symptoms Index. *Journal of Aging and Health*, *5*(2), 179–193. doi:10.1177/089826439300500202

Kunzmann, U. (2008). Differential age trajectories of positive and negative affect: further evidence from the Berlin Aging Study. *The journals of gerontology. Series B, Psychological sciences and social sciences*, *63*(5), P261–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/18818440

Kunzmann, U., Little, T. D., & Smith, J. (2000). Is age-related stability of subjective well-being a paradox? Cross-sectional and longitudional evidence from the Berlin Aging Study. *Psychology and Aging*, *15*(3), 511–526. doi:10.1037//0882-7974.15.3.511

Lasch, C. (1979). *The Culture of Narcissism: American Life in an Age of Diminishing Expectations.* New York: Warner.

Laslett, P. (1989). *A fresh map of life: the emergence of the Third Age* (p. 213). London: Weidenfeld and Nicolson. Retrieved from http://books.google.com/books?hl=en&lr=&id=TxqhSpbjVGgC&pgis=1

Lewinsohn, P. M., Seeley, J. R., Roberts, R. E., & Allen, N. B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and aging*, *12*(2), 277–87. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9189988

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology*, *84*(3), 527–539. doi:10.1037/0022-3514.84.3.527

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2004). Unemployment alters the set point for life satisfaction. *Psychological Science*, *15*(1), 8. Retrieved from http://pss.sagepub.com/content/15/1/8.short

Marmot, M., Banks, J., Blundell, R., Erens, B., Lessof, C., Nazroo, J., & Huppert, F. A. (2011). English Longitudinal Study of Ageing: Wave 0 (1998,1999 and 2001) and Waves 1-4 (2002-2009). Colchester, Essex: UK Data Archive.

Marsh, H. W. (1996). Positive and negative global self-esteem: a substantively meaningful distinction or artifactors? *Journal of personality and social psychology*, *70*(4), 810–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8636900

Marsh, H. W., Muthén, B. O., Asparouhov, T., Luedtke, O., Robitzch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA.. Application to Students' Evaluations of University Teaching. *Structural Equation Modeling*, *16*(April 2012), 37–41. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Structural+Equation+Modeling+:+A+Exploratory+Structural+Equation+Modeling+,+Integrating+CFA+and+EFA+:+Application+to+Students+'+Evaluations+of+University+Teaching#2

McDonald, R. P. (1999). *Test Thory: A unified treatment.* Mahwah, NJ: Lawrence Erlbaum.

Morrison, M., Tay, L., & Diener, E. (2011). Subjective well-being and national satisfaction: findings from a worldwide survey. *Psychological science*, *22*(2), 166–71. doi:10.1177/0956797610396224

Muthén, L. K., & Muthén, B. O. (2010). *Mplus User ' s Guide* (Sixth edit.). Los Angeles, CA: Muthen & Muthen.

Nave, C. S., Sherman, R. A., & Funder, D. C. (2008). Beyond Self-Report in the Study of Hedonic and Eudaimonic Well-Being: Correlations with Acquaintance Reports, Clinician Judgments and Directly Observed Social Behavior. *Journal of research in personality*, *42*(3), 643–659. doi:10.1016/j.jrp.2007.09.001

Nolan, J. L. J. (1998). *The Therapeutic State: Justifying Government at Century's End.* New York: New York University Press.

Nussbaum, M., & Sen, A. (1993). *The Quality of Life*. (M. C. Nussbaum & A. K. Sen, Eds.)*Development* (Vol. 1, p. xi, 453). Oxford University Press. doi:10.1093/0198287976.001.0001

Oishi, S. (2006). The concept of life satisfaction across cultures: An IRT analysis. *Journal of Research in Personality*, *40*(4), 411–423. doi:10.1016/j.jrp.2005.02.002

Papassotiropoulos, A., Heun, R., & Maier, W. (1999). The impact of dementia on the detection of depression in elderly subjects from the general population. *Psychological medicine*, *29*(1), 113–20. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10077299

Parmelee, P. A. (2007). Depression. *Encyclopedia of Gerontology*.

Pavot, W., & Diener, E. (1993). Review of the satisfaction with life scale. *Psychological Assessment*, *5*(2), 164. Retrieved from http://psycnet.apa.org/journals/pas/5/2/164/

Pons, D., Atienza, F. L., Balaguer, I., & García-Merita, M. L. (2000). Satisfaction with life scale: analysis of factorial invariance for adolescents and elderly persons. *Perceptual and motor skills*, *91*(1), 62–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11011872

Radloff, L. S. (1977). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied psychological measurement*, *1*(3), 285–401.

Raju, N. S., Laffitte, L. J., & Byrne, B. M. (2002). Measurement equivalence: A comparison of methods based on confirmatory factor analysis and item response theory. *Journal of Applied Psychology*, *87*(3), 517–529. doi:10.1037//0021-9010.87.3.517

Ready, R. E., Vaidya, J. G., Watson, D., Latzman, R. D., Koffel, E. a, & Clark, L. A. (2011). Age-group differences in facets of positive and negative affect. *Aging & mental health*, *15*(6), 784–95. doi:10.1080/13607863.2011.562184

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory Factor Analysis and Item Response Theory: Two Approaches for Exploring Measurement Invariance. *Psychological Bulletin*, *114*(3), 552–566.

Ross, C. E., & Mirowsky, J. (1984). Components of depressed mood in married men and women. *American Journal of epidemiology*, *119*(6), 997–1004.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American psychologist*, *55*(1), 68–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11392867

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology*, *52*(1), 141–166. Retrieved from http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.52.1.141

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of personality and social psychology*, *69*(4), 719. Retrieved from http://psycnet.apa.org/psycinfo/1996-08070-001

Ryff, C. D., & Singer, B. H. (1998). The contours of positive human health. *Psychological inquiry*, *9*(1), 1–28. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15327965pli0901_1

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American psychologist*, *55*(1), 5. Retrieved from http://psycnet.apa.org/journals/amp/55/1/5/

Shevlin, M., & Adamson, G. (2005). Alternative factor models and factorial invariance of the GHQ-12: a large sample analysis using confirmatory factor analysis. *Psychological assessment*, *17*(2), 231–6. doi:10.1037/1040-3590.17.2.231

Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of consumer research*, *25*(1). Retrieved from http://www.jstor.org/stable/10.1086/209528

Steptoe, A., & Wardle, J. (2011). Positive affect measured using ecological momentary assessment and survival in older men and women. *Proceedings of the National Academy of Sciences*, *108*(45), 18244–18248. doi:10.1073/pnas.1110892108

Sumner, L. (1999). *Welfare, happiness, and ethics*. Oxford: Oxford University Press.

Szasz, T. S. (1999). The Therapeutic State. *Spring*, 485–521.

Taylor, S. E., & Brown, J. D. (1988). Illusion and Well-being: A Social Psychological Perspective on Mental Health. *Psychological Bulletin*, *103*(2), 193–210. Retrieved from

http://www.lrsi.uqam.ca/documents/PSY9520/05 - l%27estime de soi 2 - ses fonctions, cons%E9quences, et processus alternatifs/TAYLOR~1.PDF

Van de Velde, S., Bracke, P., Levecque, K., & Meuleman, B. (2010). Gender differences in depression in 25 European countries after eliminating measurement bias in the CES-D 8. *Social Science Research*, *39*(3), 396–404. doi:10.1016/j.ssresearch.2010.01.002

Van den Berg, M. D., Oldehinkel, a J., Bouhuys, a L., Brilman, E. I., Beekman, a T., & Ormel, J. (2001). Depression in later life: three etiologically different subgroups. *Journal of affective disorders*, *65*(1), 19–26. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11426505

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, *3*(1), 4–70. doi:10.1177/109442810031002

Wallace, R. B., Herzog, A. R., Ofstedal, M. B., Steffick, D., Fonda, S., & Langa, K. M. (2000). *Documentation of Affective Functioning Measures in the Health and Retirement Study*. Ann Arbor, MI.

Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, *64*(4), 678–691. Retrieved from http://psycnet.apa.org/journals/psp/64/4/678/

Waterman, A. S. (2007). On the importance of distinguishing hedonia and eudaimonia when contemplating the hedonic treadmill. *The American psychologist*, *62*(6), 612–3. doi:10.1037/0003-066X62.6.612

Watson, D., Clark, L. A., & Tellegen, a. (1988). Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology*, *54*(6), 1063–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3397865

Watson, L. C., & Pignone, M. P. (2003). Screening accuracy for late-life depression in primary care: A systematic review. *Journal of Family Practice*, *52*(12), 956–64. Retrieved from http://www.jfponline.com/pages.asp?aid=1596

Wiggins, R. D., Netuveli, G., Hyde, M., Higgs, P., & Blane, D. (2007). The Evaluation of a Self-enumerated Scale of Quality of Life (CASP-19) in the Context of Research on Ageing: A Combination of Exploratory and Confirmatory Approaches. *Social Indicators Research*, *89*(1), 61–77. doi:10.1007/s11205-007-9220-5

Wood, A. M., Taylor, P. J., & Joseph, S. (2010). Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *Psychiatry Research*, *177*(1-2), 120–123. doi:10.1016/j.psychres.2010.02.003

Wu, C., & Yao, G. (2006). Analysis of factorial invariance across gender in the Taiwan version of the Satisfaction with Life Scale. *Personality and Individual Differences*, *40*(6), 1259–1268. doi:10.1016/j.paid.2005.11.012

Appendix

1. Measurement instruments

SWLS (Diener, 1984)

a. In most ways my life is close to ideal
b. The conditions of my life are excellent
c. I am satisfied with my life
d. So far, I have gotten the important things I want in life
e. If I could live my life again, I would change almost nothing

Answering categories

1: Strongly agree
2: Agree
3: Slightly agree
4: Neither agree nor disagree
5: Slightly disagree
6: Disagree
7: Strongly disagree

CES-D (Radloff, 1977)

Now think about the past week and the feelings you have experienced. Please tell me if each of the following was true for you much of the time during the past week.

(Much of the time during past week),

a. You felt depressed?
b. You felt that everything you did was an effort?
c. Your sleep was restless
d. You were happy
e. You felt lonely
f. You enjoyed life
g. You felt sad
h. You could no get going

Answering categories

1: Yes
2: No

GHQ (Goldberg, 1988)

We should like to know how your health has been in general over the past few weeks. Have you recently…

a. been able to concentrate on whatever you're doing?
b. lost much sleep over worry?
c. felt you were playing a useful part in things?
d. felt capable of making decisions?
e. felt constantly under strain?
f. felt you couldn't overcome your difficulties?
g. been able to enjoy your normal day-to-day activities?
h. been able to face up to your problems?
i. been feeling unhappy and depressed?
j. been losing confidence in yourself?
k. been thinking of yourself as a worthless person?
l. been feeling reasonably happy, all things considered?

Answering categories

1 Better than usual
2 Same as usual
3 Less than usual
4 Much less than usual

Psychological Well-being (Ryff, 1989)

Purpose in life

a. I enjoy making plans for the future and working to make them a reality.
b. My daily activities often seem trivial and unimportant to me.
c. I am an active person in carrying out plans for myself.
d. I don't have a good sense of what it is I'm trying to accomplish in life.
e. I sometimes feel as if I've done all there is in life.
f. I live life one day at a time and don't really think about the future.
g. I have a sense of direction and purpose in my life.

Personal Growth

h. I am not interested in activities that will expand my horizons.
i. I think it is important to have new experiences that challenge how I think about myself and the world.
j. When I think about it, I haven't really improved much as a person over the years.
k. I have the sense that I have developed a lot as a person over time.
l. I do not enjoy being in new situations that require me to change my old familiar ways of doing things.
m. I gave up trying to make big improvements in my life a long time ago.
n. For me, life has been a continuous process of learning, changing and growth.

Self acceptance

o. I feel like many of the people I know have gotten more out of life than I have.
p. In general, I feel confident and positive about myself.
q. When I compare myself to friends and acquaintances, it makes me feel good about who I am.
r. My attitude about myself is probably not as positive as most people feel about themselves.
s. In many ways, I feel disappointed about my achievements in life.
t. When I look at the story of my life, I am pleased with how things have turned out.
u. I like most parts of my personality.

Answering categories

1. Completely disagree
2. Disagree
3. Somehow disagree
4. Somehow agree
5. Agree
6. Completely agree

Quality of life (CASP) (Hyde et al., 2003)

Here is a list of statements that people have used to describe their lives or how they feel. We would like to know how often, if at all, you think they apply to you.

Control

a. My age prevents me from doing the things I would like to.
b. I feel that what happens to me is out of control.
c. I feel free to plan things for the future.
d. I feel left out of things.

Autonomy

e. I can do the things that I want to do.
f. Family responsibilities prevent me from doing what I want to do.
g. I feel that I can please myself what I can do.
h. My health stops me from doing the things I want to do.
i. Shortage of money stops me from doing the things I want to do.

Pleasure

j. I look forward to each day.
k. I feel that my life has meaning.
l. I enjoy the things that I do.
m. I enjoy being in the company of others.
n. On balance, I look back on my life with a sense of happiness.

Self-realization

o. I feel full of energy these days.
p. I choose to do things that I have never done before.
q. I feel satisfied with the way my life has turned out.

r. I feel that life is full of opportunities.
s. I feel that the future looks good for me.

Answering categories

1 Often
2 Sometimes
3 Not often
4 Never

2. Model specification Figures

Figure A: 1 Factor model for CASP 19

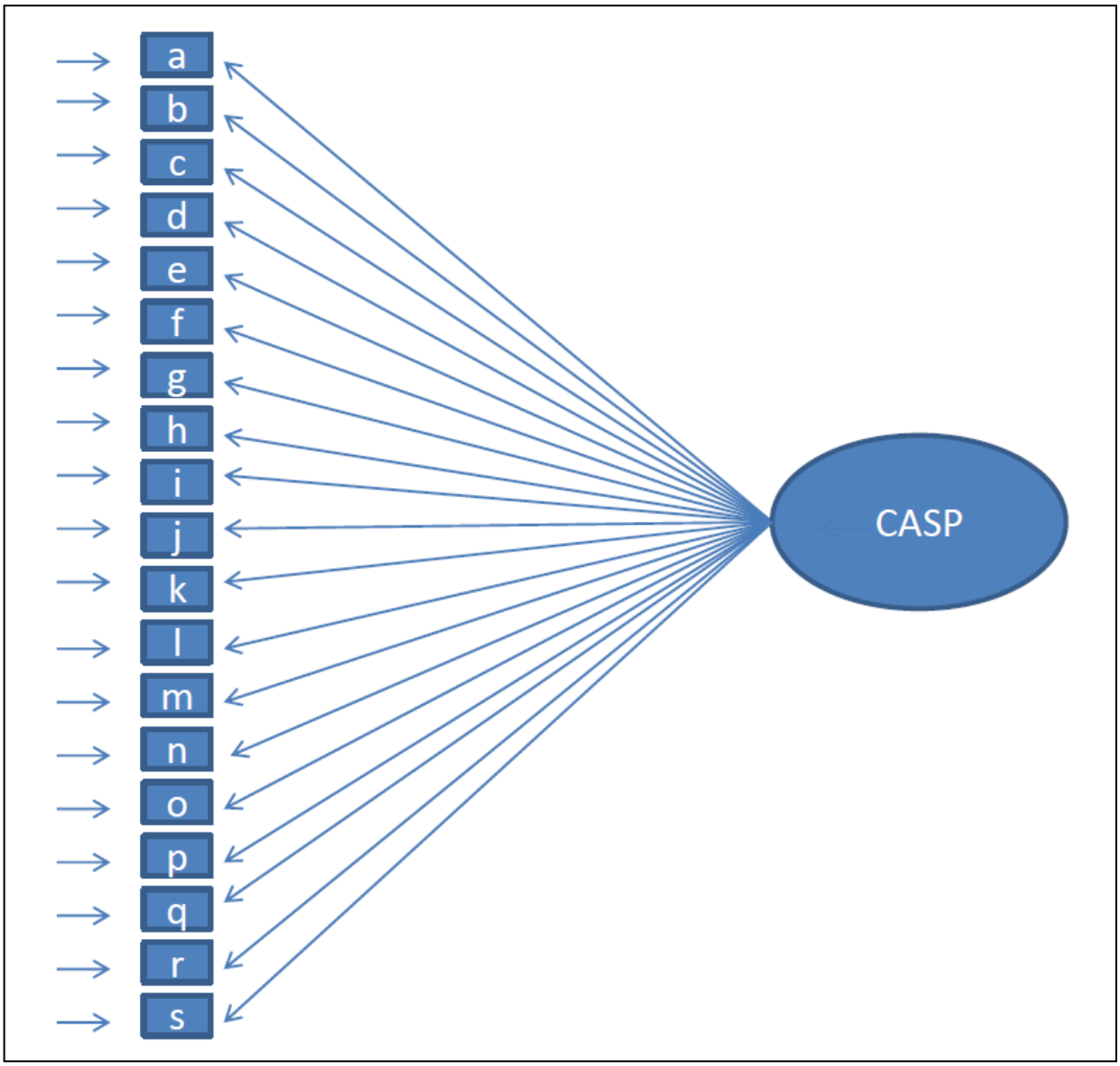


Figure B: 1 Factor model for CASP 19 with error correlations

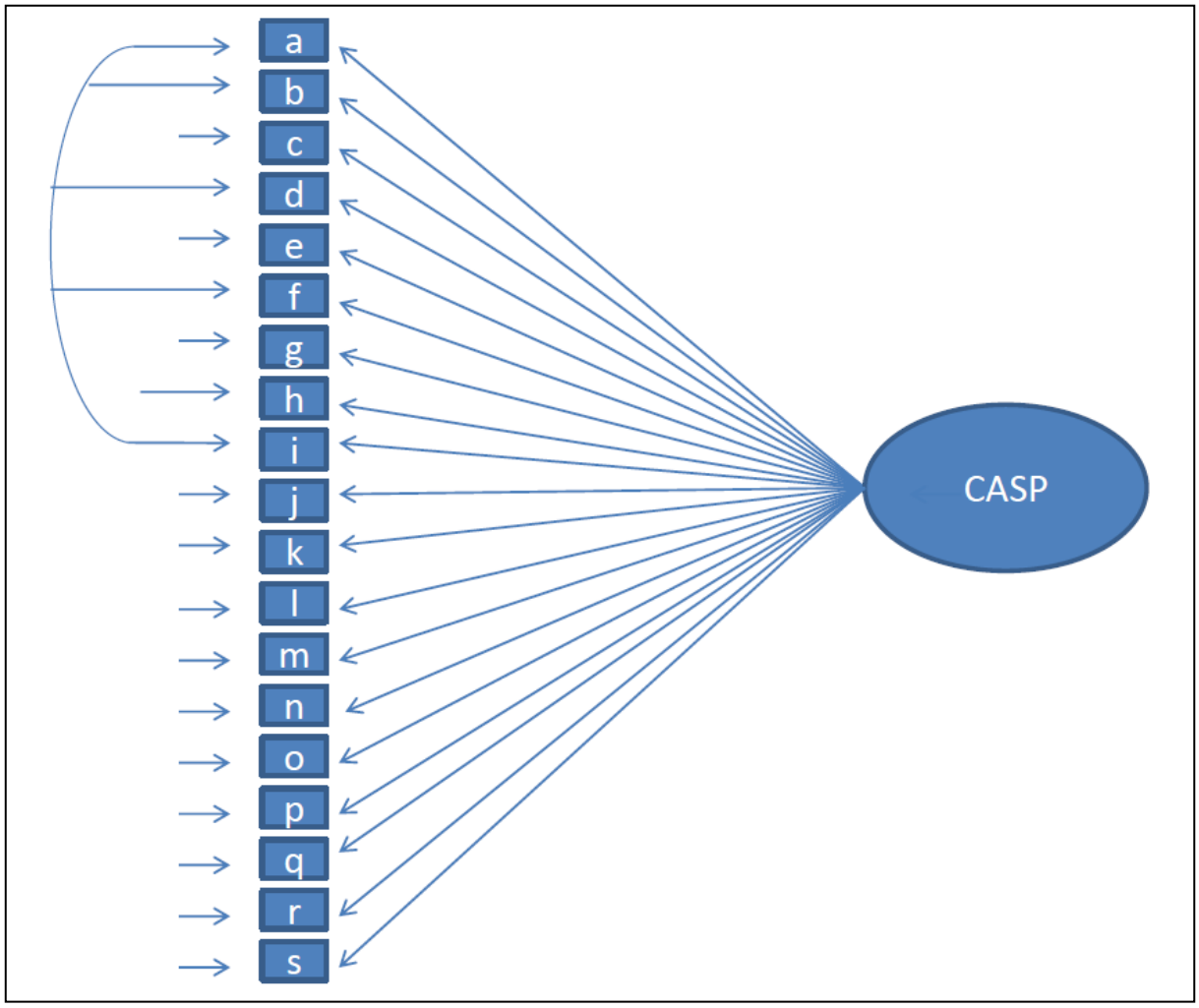


Figure C: 1 Factor model for CASP 19 with method factor for negatively worded items

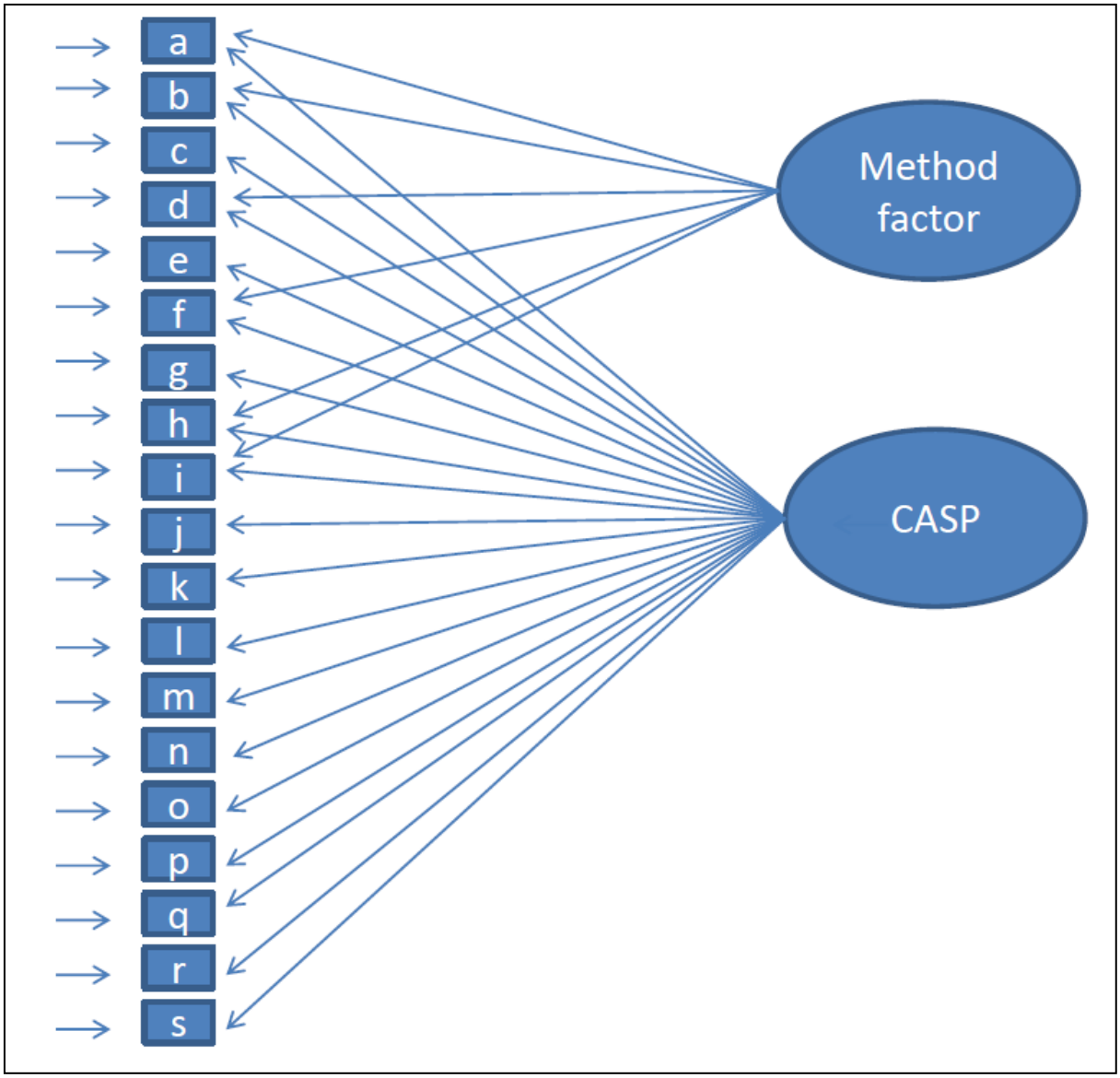


Figure D: 2 Factor model for CASP 19

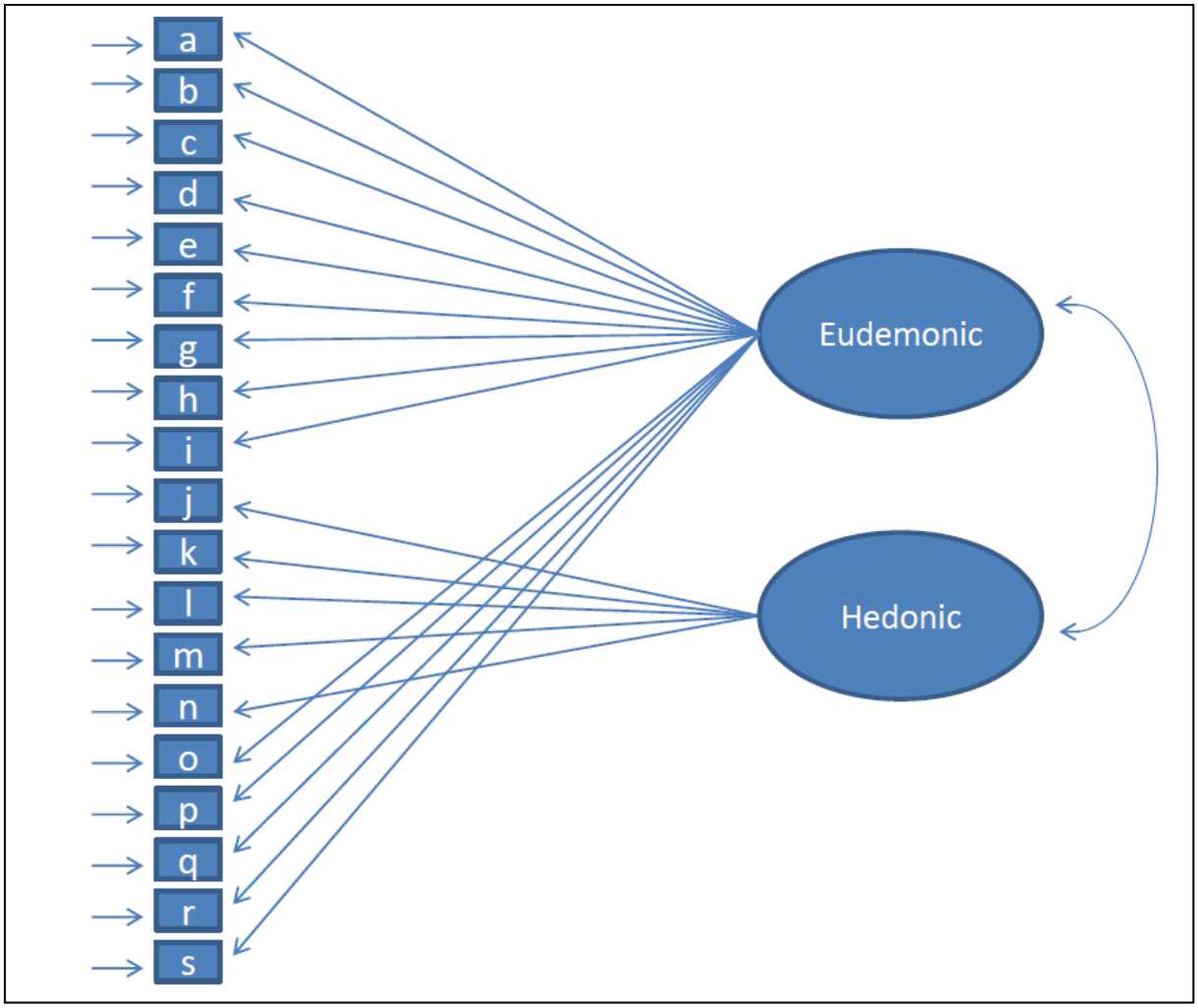
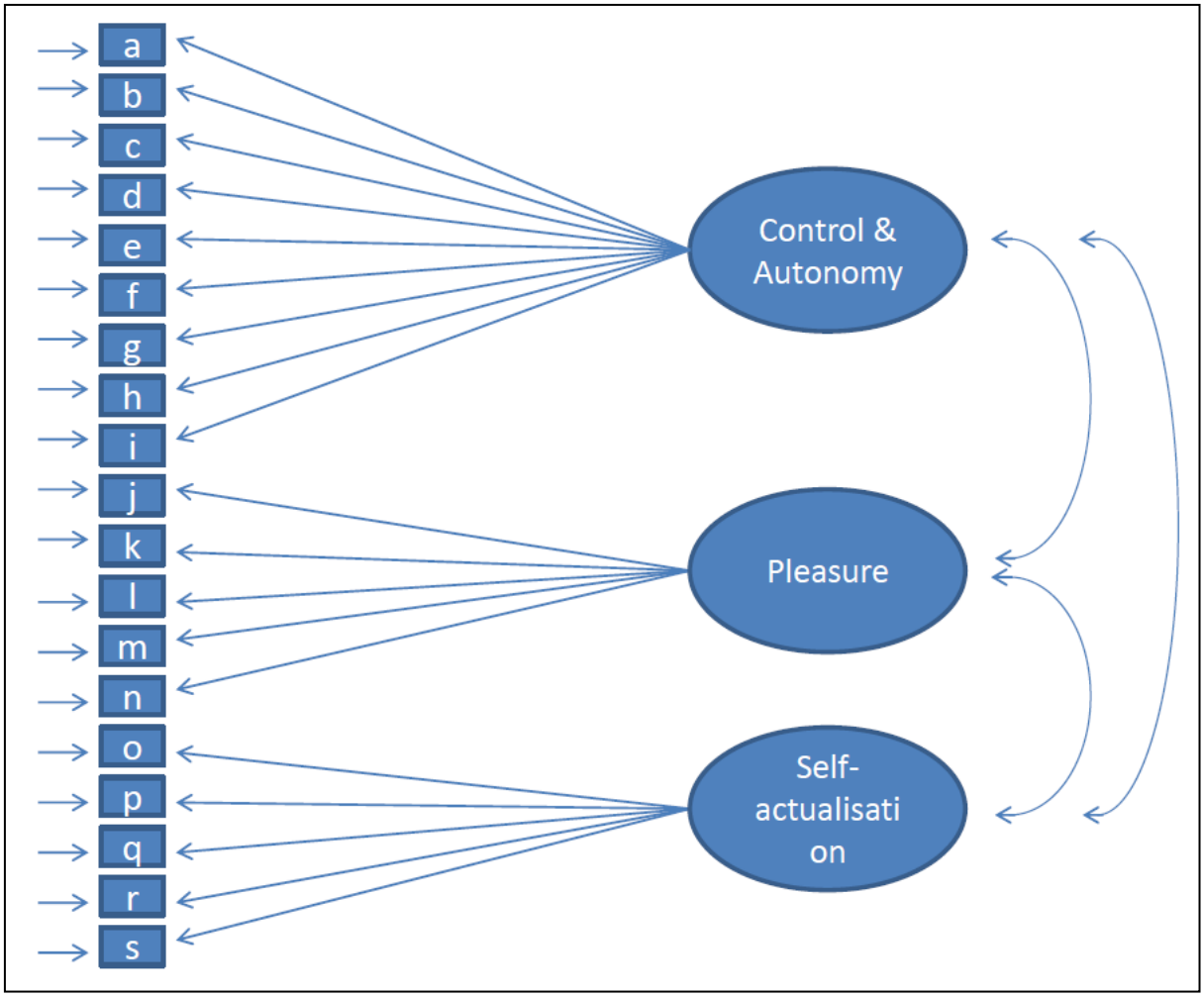


Figure E: 3 Factor model for CASP 19



3. Descriptive statistics on used data

ELSA Wave 1

CASP

Item	Variable name	n	mean	s.d.
a	scqola	10149	2.81	0.97
b	scqolb	9991	3.08	0.94
c	scqolc	9941	1.78	0.99
d	scqold	10017	3.17	0.91
e	scqole	10089	1.59	0.83
f	scqolf	10032	3.05	0.94
g	scqolg	10104	1.55	0.79
h	scqolh	10131	2.84	1.09
i	scqoli	10104	2.64	1.02
j	scqolj	10151	1.29	0.60
k	scqolk	10075	1.44	0.72
l	scqoll	10175	1.22	0.50
m	scqolm	10187	1.31	0.54
n	scqoln	10162	1.35	0.61
o	scqolo	10133	2.06	0.84
p	scqolp	10066	2.41	0.87
q	scqolq	10106	1.60	0.76
r	scqolr	10060	1.90	0.85
s	scqols	10092	1.81	0.83

CESD

Item	Variable name	n	mean	s.d.
a	psceda	11040	1.82	0.38
b	pscedb	11035	1.76	0.43
c	pscedc	11041	1.59	0.49
d	pscedd	11003	1.11	0.31
e	pscede	11039	1.86	0.35
f	pscedf	10997	1.10	0.30
g	pscedg	11035	1.79	0.41
h	pscedh	11026	1.78	0.41

GHQ

Item	Variable name	n	mean	s.d.
a	scghqa	10175	2.11	0.44
b	scghqb	10163	1.79	0.73
c	scghqc	10082	2.09	0.53
d	scghqd	10176	2.02	0.38
e	scghqe	10160	1.91	0.72
f	scghqf	10137	1.76	0.69
g	scghqg	10180	2.12	0.50
h	scghqh	10167	2.05	0.40
i	scghqi	10160	1.70	0.76
j	scghqj	10159	1.62	0.73
k	scghqk	10161	1.34	0.62
l	scghql	10169	1.99	0.45

ELSA Wave 2

SWLS

Item	Variable name	n	mean	s.d.
a	sclifea	7585	5.08	1.42
b	sclifeb	7527	5.14	1.48
c	sclifec	7654	5.53	1.34
d	sclifed	7651	5.69	1.27
e	sclifee	7661	4.80	1.78

SHARE Wave 2

CASP

Item	Variable name	n	mean	s.d.
a	ac014	33610	2.64	1.03
b	ac015	33444	2.84	0.96
d	ac016	33506	3.05	0.96
e	ac017	33563	1.77	0.88
f	ac018	33564	3.04	0.97
i	ac019	33573	2.56	1.10
j	ac020	33429	1.67	0.92
k	ac021	33368	1.44	0.72
n	ac022	33275	1.61	0.76
o	ac023	33591	1.85	0.86
r	ac024	33392	1.90	0.87
s	ac025	33181	1.92	0.88

Correlation matrix CASP items, ELSA Wave 1 (n=9300)

	scqola	scqolb	scqolc	scqold	scqole	scqolf	scqolg	scqolh	scqoli	scqolj	scqolk	scqoll	scqolm	scqoln	scqolo	scqolp	scqolq	scqolr	scqols
scqola	1.00																		
scqolb	0.42	1.00																	
scqolc	-0.18	-0.22	1.00																
scqold	0.32	0.45	-0.22	1.00															
scqole	-0.25	-0.26	0.47	-0.24	1.00														
scqolf	0.03	0.16	-0.04	0.19	-0.02	1.00													
scqolg	-0.10	-0.17	0.34	-0.16	0.43	-0.28	1.00												
scqolh	0.61	0.43	-0.23	0.35	-0.32	0.02	-0.11	1.00											
scqoli	0.15	0.22	-0.16	0.23	-0.14	0.22	-0.15	0.19	1.00										
scqolj	-0.17	-0.27	0.32	-0.29	0.34	-0.07	0.26	-0.20	-0.12	1.00									
scqolk	-0.18	-0.25	0.32	-0.29	0.29	-0.02	0.23	-0.20	-0.11	0.54	1.00								
scqoll	-0.19	-0.26	0.31	-0.29	0.36	-0.10	0.32	-0.20	-0.15	0.58	0.50	1.00							
scqolm	-0.08	-0.12	0.18	-0.13	0.19	-0.04	0.19	-0.11	-0.06	0.32	0.31	0.36	1.00						
scqoln	-0.10	-0.21	0.25	-0.26	0.24	-0.09	0.23	-0.14	-0.18	0.44	0.41	0.43	0.34	1.00					
scqolo	-0.48	-0.40	0.34	-0.33	0.40	-0.01	0.22	-0.59	-0.16	0.41	0.38	0.39	0.24	0.31	1.00				
scqolp	-0.33	-0.23	0.30	-0.20	0.30	0.02	0.20	-0.34	-0.07	0.28	0.30	0.28	0.23	0.20	0.48	1.00			
scqolq	-0.22	-0.32	0.33	-0.37	0.34	-0.09	0.25	-0.29	-0.26	0.47	0.45	0.46	0.27	0.53	0.44	0.31	1.00		
scqolr	-0.30	-0.30	0.37	-0.32	0.33	-0.04	0.25	-0.33	-0.19	0.42	0.44	0.40	0.28	0.37	0.50	0.46	0.50	1.00	
scqols	-0.35	-0.38	0.42	-0.39	0.40	-0.06	0.27	-0.40	-0.24	0.49	0.49	0.45	0.28	0.42	0.57	0.43	0.61	0.64	1.00

Correlation matrix CES-D items, ELSA Wave 1 (n=10940)

	psceda	pscedb	pscedc	pscedd	pscede	pscedf	pscedg	pscedh
psceda	1.00							
pscedb	0.47	1.00						
pscedc	0.29	0.30	1.00					
pscedd	-0.44	-0.30	-0.20	1.00				
pscede	0.35	0.28	0.17	-0.30	1.00			
pscedf	-0.41	-0.33	-0.21	0.58	-0.31	1.00		
pscedg	0.50	0.33	0.23	-0.39	0.39	-0.37	1.00	
pscedh	0.40	0.53	0.29	-0.29	0.25	-0.32	0.31	1

Correlation matrix GHQ items, ELSA Wave 1 (n=9934)

	scghqa	scghqb	scghqc	scghqd	scghqe	scghqf	scghqg	scghqh	scghqi	scghqj	scghqk	scghql
scghqa	1.00											
scghqb	0.32	1.00										
scghqc	0.40	0.23	1.00									
scghqd	0.40	0.19	0.39	1.00								
scghqe	0.34	0.55	0.24	0.23	1.00							
scghqf	0.36	0.47	0.32	0.29	0.59	1.00						
scghqg	0.46	0.32	0.46	0.35	0.38	0.42	1.00					
scghqh	0.40	0.30	0.38	0.43	0.34	0.41	0.50	1.00				
scghqi	0.37	0.53	0.32	0.27	0.57	0.56	0.39	0.40	1.00			
scghqj	0.39	0.43	0.37	0.35	0.49	0.55	0.38	0.41	0.64	1.00		
scghqk	0.33	0.34	0.37	0.31	0.39	0.45	0.33	0.40	0.52	0.63	1.00	
scghql	0.36	0.31	0.34	0.33	0.34	0.37	0.43	0.47	0.45	0.41	0.42	1.00

Correlation matrix SWLS items, ELSA Wave 2 (n=7393)

	sclifea	sclifeb	sclifec	sclifed	sclifee
sclifea	1				
sclifeb	0.77	1.00			
sclifec	0.76	0.80	1.00		
sclifed	0.64	0.63	0.72	1.00	
sclifee	0.56	0.53	0.57	0.54	1

Correlation matrix CASP items, SHARE Wave 2 (n=32258)

	ac014	ac015	ac016	ac017	ac018	ac019	ac020	ac021	ac022	ac023	ac024	ac025
ac014	1.00											
ac015	0.42	1.00										
ac016	0.39	0.53	1.00									
ac017	-0.25	-0.21	-0.25	1.00								
ac018	0.14	0.19	0.21	-0.07	1.00							
ac019	0.23	0.21	0.26	-0.18	0.28	1.00						
ac020	-0.14	-0.15	-0.21	0.23	-0.06	-0.11	1.00					
ac021	-0.25	-0.25	-0.32	0.33	-0.06	-0.17	0.42	1.00				
ac022	-0.18	-0.18	-0.24	0.23	-0.09	-0.20	0.27	0.44	1.00			
ac023	-0.40	-0.34	-0.34	0.36	-0.06	-0.17	0.27	0.44	0.33	1.00		
ac024	-0.34	-0.27	-0.32	0.37	-0.08	-0.25	0.28	0.46	0.38	0.56	1.00	
ac025	-0.36	-0.31	-0.35	0.37	-0.09	-0.28	0.31	0.49	0.41	0.54	0.63	1.00