ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA
SCUOLA DI PSICOLOGIA E SCIENZE DELLA FORMAZIONE

CORSO DI LAUREA MAGISTRALE IN
Psicologia Clinica

# Measurement Invariance of the Warwick-Edinburgh Mental Well-Being Scale and its Short form across UK and Italy

Tesi di laurea in
Tecniche di Valutazione Testistica in Psicologia Clinica

Relatore
Prof. Paola Gremigni

Presentata da
Claudia Lega

Sessione III

Anno Accademico 2012/2013

# Table of Contents

# Chapter 1. Introduction

## 1.1. The study of well-being

### 1.1.1. The concepts of mental health and well-being

The study of well-being characterized the last 25 years of research in positive psychology.
Questions about what the concept of happiness means and what make individuals happy have been debated for millennia, and answers have changed through ages according to cultures and historical moments. As far as the psychology field is concerned, the concept of well-being is included within the areas of mental health and positive psychology.
Health is one of the most important values in life, and it is recognized as a form of human capital (Keyes, 2013).
The word "health" takes its origins from the Old English term *hale*, meaning "wholeness", that in turn derives from the Proto-Indo-European root *kailo*, that stands for "whole, uninjured".
In the 1950s, the World Health Organization recognized the importance of focusing not only on the diagnosis and the cure of the illness but also on the positive aspects of health. From this outlook arose the definition of WHO (1947), describing health as "*a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity*" (WHO, 1947, p.1). Mental health, including cognitive and emotional competencies, is an indispensable component of health, that is why HM Government (2011) stated that there is no health without mental health.
In 2001, The World Health Organization has defined mental health as "*a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community*" (WHO, 2001, p.1).
This perspective emerges in the study of positive psychology, a recent discipline developed at the end of the 1990s centered on the good components of human functioning.
This branch explores people's strengths and virtues in order to increase their fulfilment and pleasant feelings. The aim of positive psychology is to shift the emphasis from purely healing the sickness to strengthening good abilities (Seligman & Csikszentmihalyi, 2000).
The concept of well-being appears in WHO's definitions of both health and mental health, calling attention to the assessment of good side of functioning such has human potentialities, good qualities, and general sense of happiness.
Well-being has been a paramount concern of thinkers since ancient times, and it became a topic of scientific inquiry during the 1950s when, after the World War II, there was a great

interest in social welfare, appreciation of the individual, and importance of personal meaning and concerns about life. Social scientists developed indicators of quality of life to monitor social change and to improve social policy. During the same historical period the National Institute of Mental Health (NIMH) was funded and although its aim was to promote the America's mental health, its main work was to study the etiology and treatment of mental illness.

The study of subjective well-being has been traditionally divided into two streams of research. The first conception is called hedonic tradition (from the Greek word *Ἡδονή*, hedoné, that means pleasure) and equates well-being with happiness as feeling good. It reflects the Epicurean view that happiness is about feeling positive emotions and avoiding negative ones.

The second conception is called eudaimonic tradition, from the Greek words *εὖ*, eu (good), and δαίμων, daimon (spirit). It equates well-being with the human potential of pursuing and developing a positive functioning in life, and it reflects the Aristotelian and Socratic view that happiness is about striving toward excellence and good performance as a member of society. This second tradition is reflected in the stream of research on subjective psychological well-being (Ryff, 1989), and subjective social well-being (Keyes, 1998).

According to the model of Carol Ryff (1989), positive functioning consists of six dimensions: self-acceptance, positive relations with others, personal growth, purpose in life, environmental mastery, and autonomy. This six-factor structure has been confirmed in the national survey MIDUS (Midlife Development in the U.S, Ryff &Keyes, 1995) and from this model the Psychological Well-Being Scales have been developed (Ryff, 1989; Italian validation by Ruini et al., 2003).

On the other side, social well-being (Keyes, 1998) defines a more public experience focused on the social tasks that individuals meet in interpersonal contexts. Social well-being consists of five elements (social integration, social contribution, social coherence, social actualization, social acceptance) that evaluate whether and to which extent people are functioning well in their social world.

### 1.1.2. The importance of studying well-being

The benefits of well-being are significant at different levels.

Lyubormirsky, King & Diener (2005) reviewed three classes of studies (cross-sectional, longitudinal and experimental) to understand the benefits of happiness and positive affect.

Cross-sectional evidence showed that happy people are more successful in work, relationships and health, and that long-term well-being and short-term positive affect are associated with positive understanding of self and others, pro-social behaviour, healthy behaviour, high

immune functioning, and good coping distress. Longitudinal studies revealed that happiness precedes fulfilling work, satisfying relationships, better mental and physical health, and longevity. Experimental researches demonstrated that even temporary happy moods make people engage more with the environment and be more venturesome, open and sensitive.

Diener and Chan (2011) examined seven types of evidence (including longitudinal studies) discovering that high subjective well-being (SWB) causes better health and longevity in many nations, even if it is still controversial whether SWB could improve chances of surviving illnesses. Howell and colleagues (2007) not only showed that SWB is related to both cardiovascular health and better immune functioning, but they also estimated a 14% longevity difference between happy and unhappy individuals. Social relationships have a larger effect on longevity than factors like physical activity, body-mass index and air pollution (Holt-Lunstand, Smith & Layton, 2010).

Since the moment psychology finally recognized the priority of focusing on well-being and its components, lot of work has been done in this field. Different instruments has been developed to evaluate these aspects and many of them showed a good performance and applicability. However, as I will describe hereinafter, there is a necessity of having precise and convincing tools which must show not only good psychometric properties of validity and reliability, but which also are applicable in a more extended context. Nowadays in fact, numerous studies are interested in evaluating the differences in many situations, and researches are always more frequently conducted on more than one sample of subjects. For this reason, the concept of Measurement Invariance, that indicate the quality possessed by an instrument to maintain the same psychometric properties even in different samples or under different conditions, is considered fundamental in the field of cross-cultural studies.

This is particularly true in the developing area of well-being. Indeed, while lots of well-known instruments to detect symptoms of illness or problems in the individuals have been tested many times and modified in order to perform better, this still has to happen with many tools currently used. As I will illustrate more deeply in this work, WEMWBS is a young tool (developed in 2007) which quickly become widespread thanks to its good characteristics of validity and its easiness of use. It is nowadays used even in the National Survey for England (NSE) that represent the most accurate research of health trends in general population. WEMWBS was also validated in different countries (Stewart-Brown, 2013), but its invariance across nationalities has never been verified.

## 1.2. International and individual differences in well-being

### 1.2.1. The role of country

Diener et al. (1995) investigated subjective well-being (SWB) in 55 nations. They evaluated 6 predictor variables: wealth, rights, growth of wealth, income social comparison, equality, and individualism-collectivism. The findings revealed that SWB was correlated with social, economic, and cultural characteristics of the nations. High income, individualism, human rights, and societal equality correlated strongly with each other and with SWB. Income correlated with SWB even after basic need fulfilment was controlled. This indicates that economic development has an impact on well-being that transcend meeting basic biological needs such as food, water, health and sanitation. Cultural homogeneity, income growth, and income comparison showed either low or inconsistent relations with SWB. It has been noted that Greece, France, Italy and Spain report lower levels of SWB than one might expect based on their GDP (Gross Domestic Product) per person. Value of Mean SWB for Italy was -0.44, while in Britain it was 0.69.

Layard, Mayraz and Nickell (2010) showed how in advanced countries such as USA and Germany levels of happiness and life satisfaction have not risen since 1950, despite the increasing economic growth. Interestingly, this seems explained more by the effect of relative income (i.e. the comparison with other people income), rather than the absolute income. In the same study the authors investigated the long-run effect of higher living standards in 16 Western Europe Countries (including Italy and United Kingdom), covering the period from 1973 to 2007. Again, the income of the whole population loaded less on average life satisfaction than the increase in just one's person income, remarking the role of social comparison. Reported levels of life satisfaction (measured with the Eurobarometer) were lower in Italy than in United Kingdom.

Ferrara and Nisticò (2013) examined well-being indicators in Italian regions by developing two composite indicators, the Augmented Human Development Index (AHDI) and the Well-Being Index (WBI). The AHDI was obtained by combining three indicators (health, adult education, per-capita disposable income) inspired by a previous study of Marchante (2006), while the WBI considered the previous dimensions plus gender equal-opportunities, market abilities and the quality of the socio-institutional context. The assessment of well-being based on the current situation shows a clear separation between Centre-Northern and Southern regions with the first reporting higher levels of quality of life, less age and gender discrimination and better quality of the context compared to the latter. However, Italian regions tended to become more similar in terms of well-being over the 1998-2008 decade,

when one Southern region (Abruzzo) had a higher value of well-being than the national average, and that Southern regions presented WBI values of 63%.

Brown et al. (2008) argued that well-being is influenced not just by relative income (as already shown by different authors such as Clarke et al., 2008; Sacks et al.2010), but also by the rank-ordered position of the individual wage within a comparison set. Using three complementary methods (a laboratory-based study, large-scale surveys and an analysis of quits), they showed the importance of comparative relations with a reference group. This is also in line with the study of McBride (2001), who emphasized the importance of a reference group in determining job satisfaction.

### 1.2.2. The role of gender

It is well-established that women earn less than men (Oostendorp, 2009; Albanesi & Olivetti, 2009; Weichselbaumer & Winter-Ebmer, 2005). Nevertheless, as Lalive and Stutzer (2010) states, women do not report significantly lower satisfaction than men with their life and job. These authors also found that women in conservative areas with lack of equality and a large gender wage gap are were more satisfied with life than men.

Abbot and colleagues (2008) conducted a birth Cohort study with a sample of 1134 British women to analyse the effect of individual differences in personality traits (in particular, extroversion and neuroticism) on PWB. More extroverted women reported higher well-being on all six dimensions of Carol Ryff model, while neuroticism was strongly associated with three PWB dimensions (environmental mastery, purpose in life and self-acceptance). However, the structural equation models used to examine the mediation of personality traits among latent variables, showed that the effect of early neuroticism was almost entirely mediated through emotional adjustment.

Schmitt, Branscombe, Kobrynowicz, and Owen (2002) examined the effects of gender discrimination on psychological well-being. Women perceived more discrimination than men, and they partially coped with the negative consequences by increasing the identification with their own gender (women as a group). In contrast, perceived discrimination was unrelated to group identification among men. According to the authors, this difference was due to groups' relative positions within the social structure.

Kim and Moen (2002) investigated the relationship between retirement transitions and subsequent psychological well-being in 458 people aged 50-72 years. Statistically significant gender differences in well-being measures were found. Men were higher than women in morale, personal control and income adequacy, while women reported more depressive symptoms. However, the work or retirement status of the individual and of his spouse was

more strongly related to men's psychological well-being than women. This shows that gender is a key factor in the transition to retirement.

Nolen and Rustin (2003) reviewed gender difference in major symptoms of psychopathology and found consistent gender differences in moods (sadness, anxiety, fear) and behaviors (antisocial personality disorder, conduct disorder, substance abuse or dependency). These two categories were more experienced by women, while men reported more externalizing disorder. However, women also reported a greater expression of positive moods than men. Explanations for these differences referred to biological traits (attributed to hormonal influences and genetic predispositions), personality (Type A) and social context (status, role and expectations).

### 1.2.3. The role of age

Manzi and Vignoli (2006) examined the role of family in a sample of 24 Italian and 109 U.K. adolescents. Confirmatory factor analyses showed that cohesion and enmeshment were distinguishable in both countries, orthogonal in the U.K. but positively correlated in Italy. Family cohesion was associated with better psychological well-being in both countries; enmeshment was associated with poorer psychological well-being in the U.K. but not in Italy. . Items were completely independent dimensions among the U.K. sample, while they showed positive correlation in the Italian one (Italian adolescents describing their families as more cohesive tended to describe them also as more enmeshed). While psychological well-being in U.K. sample was predicted positively by family cohesion and negatively by enmeshment (this is in line with previous studies on English families showing a strong cultural emphasis on autonomy, individuation, psychological distance and physical separation of young adults), in the Italian sample family enmeshment showed no relationship – positive or negative – with any well-being measures (it seems therefore that family enmeshment does not appear to be maladaptive in a traditional cultural context that emphasizes family connectedness). It seems therefore that family enmeshment (that negatively influences identity in the U.K. but not in Italy) has different overall implication for psychological well-being in the two cultures.

During 2002 and 2003, the European Union conducted the European Study on Adult Well-being (ESAW) in 6 countries (Italy, United Kingdom, Austria, Luxembourg Netherlands, and Sweden) in order to identify the factors contributing to life satisfaction for older people. The Netherlands, United Kingdom, Luxembourg and Austria had higher values in all chosen indicators of life satisfaction (i.e., physical health and functional status, self-resources, material security, social support resources, life activity) compared to Sweden and Italy. Italy was the country with the lowest fertility rate, highest unemployment rate and an early age of

retirement. Concerning the UK, it could be assumed that good standards of living underlying the satisfaction ratings may have changed since welfare reforms initiated in the 1970s, compared with other European nations, and may therefore explain the profile of satisfaction ratings obtained for the UK sample.

Regarding the gender differences, in general women showed a significantly lower satisfaction with health status and with material security. They also scored lower in subjective health and general life satisfaction. This result suggests that differences in life style and health behaviours in men and women become more pronounced in old age.

Regarding the differences due to the age group, people aged from 80 to 90 years scored lower than all the other age groups.

Gestorf and colleagues (2010) examined the evolution of levels of well-being in United Kingdom. Using long-term longitudinal data, they showed how the trend of well-being was relatively stable over time but declined rapidly in the period between 3 or 5 years prior to death, with mortality related mechanisms (e.g. deteriorating health) becoming the first drivers of late-life decline in well-being. This confirmed the terminal decline hypothesis, according to which there is a pre-terminal phase of relative stability in well-being and a terminal phase of steep, proximate-to-death decline.

### 1.3.The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)

#### 1.3.1. Description of the tool

The WEMWBS (see Table 1) is a 14 item scale designed to measure mental wellbeing. It includes both hedonic elements (such as happiness, joy, contentment) and eudaimonic elements (autonomy, positive relationships with others, purpose in life) (Taggart et al., 2013).

The scale consists of 14 positively worded items all covering both hedonic and eudemonic aspects of mental well-being: optimism, sense of usefulness, relax, interest in other people, energy, problem dealing, clear thinking, feeling good, feeling close to other people, confidence, taking decision, love, interest in new things, cheerfulness.

Individual are asked to tick the box that best describes their statement over the past two weeks, using a 5-point Likert scale with the following references:

1- None of the time

2- Rarely

3- Some of the time

4- Often

5- All of the time

All items are scored positively, so that the minimum is 14 and the maximum is 70.

***Table 1. English and Italian items of WEMWBS and SWEMWBS.***

| Item of WEMWBS | Item of SWEMWBS | English version | Italian version |
|---|---|---|---|
| 1 | 1 | I've been feeling optimistic about the future | Mi sono sentito ottimista riguardo al futuro |
| 2 | 2 | I've been feeling useful | Mi sono sentito utile |
| 3 | 3 | I've been feeling relaxed | Mi sono sentito rilassato |
| 4 | | I've been feeling interested in other people | Mi sono sentito interessato ad altre persone |
| 5 | | I've had energy to spare | Ho avuto grinta da vendere |
| 6 | 4 | I've been dealing with problems well | Ho affrontato bene i problemi |
| 7 | 5 | I've been thinking clearly | Ho pensato in modo chiaro |
| 8 | | I've been feeling good about myself | Mi sono sentito bene con me stesso |
| 9 | 6 | I've been feeling close to other people | Mi sono sentito vicino ad altre persone |
| 10 | | I've been feeling confident | Mi sono sentito sicuro di me |
| 11 | 7 | I've been able to make up my own mind about things | Sono stato in grado di prendere decisioni |
| 12 | | I've been feeling loved | Mi sono sentito amato |
| 13 | | I've been interested in new things | Mi sono interessato a cose nuove |
| 14 | | I've been feeling cheerful | Mi sono sentito di buon umore |

This scale captures a wide conception of well-being, including affective and emotional aspects, cognitive dimension and psychological functioning (Tennant et al., 2007).

The main purpose of the WEMWBS is to provide a short instrument which is easy to understand by general population, practical, inexpensive and to be included in large-scale health surveys and evaluations (Stewart-Brown, 2013).

WEMWBS is focused on the positive. It has good face validity among the general population, public health practitioners, policy makers and teenage students. It has a normal distribution in the general population with small ceiling and floor effects.

This tool has been found easy to complete, clear and unambiguous in research conducted with adult focus groups ( Tennant, Fishwick, Platt, Joseph, & Stewart-Brown, 2006) and has proved popular with practitioners and policy makers both in the UK and further afield (Tennant et al., 2007).

In Scotland, WEMWBS is now one of seven health targets for the Scottish Government. In England WEMWBS is recommended for monitoring mental well-being at national level as part of the new Public Health Outcomes Framework (Department of Health, 2012).

WEMWBS was validated also in Australia, Canada, the United States, Italy, Spain, Germany, France, Netherlands, Belgium, Iceland, India, Pakistan, Malaysia, and South Africa. Two of these research groups (Italy and South Africa) have completed formal quantitative evaluation of WEMWBS, and in this way they add much to the topic due to their valuable translations of WEMWBS into other languages.

### 1.3.2. Development and validation in the United Kingdom

In 2007, Tennant, Hiller, Fishwick, Platt, Joseph, Weich, Parkinson, Secker, and Stewart-Brown conducted the statistical analyses for the validation of the tool. The starting point for the development of the WEMWBS was the Affectometer 2 (Kammann & Flett, 1983), a scale developed in New Zealand to measure well-being who covers both eudemonic and hedonic aspects of mental health and has a good range of positive items. Initial scale testing was carried out using data collected from two samples. The first sample was composed by undergraduate and postgraduate students at Warwick and Edinburgh universities. They were asked to complete WEMWBS and between two and four other scales which were assigned randomly.

To assess the scale's test-retest reliability, a random sub-sample of the same students had to complete the WEMWBS one week later, using an identifier to match the data.

A second set of data from two representative Scottish population datasets, the Health Education Population Survey (NHS Health Scotland, 2007) and the "Well? What do you think?" Survey (Davidson, Myant, & O'Connor, 2007) was used to test the results of the student sample and the scale's capacity to discriminate between population groups.

For the statistical test only data where WEMWBS was fully completed were used, while unweighted data were used for the population sample.

Eight additional scales were included in the student sample questionnaire and one was available in the population sample. These scales were chosen to measure either same or similar concepts to WEMWBS, and they were the Positive and Negative Affect Schedule (PANAS) (Watson, Clarke, & Tellegen, 1988), the Short Depression-Happiness Scale (SDHS) (Joseph, Linley, Harwood, Lewis, & McCollam, 2004), the World Health Organization- Five Well-Being Index (WHO-5) (Heun, Bonsignore, Barkow, & Jessen, 2001), the Satisfaction With Life Scale (SWLS) (Diener, Emmons, Larsen, & Griffin, 1985), the Emotional intelligence Scale (EIS) (Schutte et al., 1998), and the Euro Quality Of Life Health Status Visual Analogue Scale (EQ-5D VAS) (Brooks, Rabin, & De Charro, 2003). Mental ill-health was evaluated in both populations with the General Health Questionnaire (GHQ) (Goldberg and Williams, 1988), while social desirability bias was assessed using the

Balanced Inventory of Desirable Response (BIDR) (Paulhus and Reid, 1991).Other variables of interest were collected in both samples: sex, age, housing tenure, self-perceived health status and employment status.

Although the UK validation of this tool reported some good psychometrics characteristics (such as a good face validity and a favourable construct validity with comparable scales), the scale also has some limitations, like its very high level of internal consistency (r = 0.94), the high susceptibility to social desirability bias, and its lengths (20 statements and 20 adjectives) (Tennant et al., 2007a).

At the beginning nine focus group were conducted, three in England and six in Scotland, for a total of 56 people. Participants were asked to complete the Affectometer 2 and to discuss their concept of positive mental health. Then a content analysis was used to identify concepts relating to mental well-being which participants thought should be included in the scale (Tennant et al., 2006). An expert panel with experts from different disciplines (psychology, psychiatric, public health, social science and health promotion) was organized to consider the analysis of focus group discussions and to agree the key concepts of mental well-being to be covered by the new scale and the wording of the new items (Tennant, Joseph, & Stewart-Brown, 2007b).

In 2012, Maheswaran and colleagues run a study to evaluate the responsiveness of the WEMWBS at both the individual and group level. The tool was used as an outcome measure in twelve different studies undertaken in different population, and three indexes (Standardised Response Mean, SRM, probability of change statistic, P, and standard error of measurement, SEM) were utilized to decide whether WEMWBS detected statistically important changes. The study demonstrated that WEMWBS is responsive to change in a variety of setting, such as schools, community settings and psychiatric hospitals, and at both investigated levels, probably because it evaluates mental well-being across both eudemonic and hedonic dimensions, and is therefore more able to detect changes.

The internal construct validity of WEMWBS was evaluated by Stewart-Brown and collaborators (2009) using a Rasch Analysis with a study in which the model was applied to data of 779 respondents of the Scottish Health Education Population Survey. As the initial fit to model was poor, some items (such as “I’ve been feeling good about myself” and “I’ve been interested in new things”) were deleted, and this led to a short version of the tool, a seven item scale called SWEMWBS (Short Warwick Edinburgh Mental Well-Being Scale). Thanks to its more robust measurement properties and brevity, SWEMWBS seems preferable to WEMWBS for monitoring mental well-being in populations, although it presents a more

restricted view of mental well-being as most of its items express aspects only of eudemonic well-being, and not hedonic well-being.

Rasch modeling was developed to investigate the psychometric properties of scales and instruments. The initial analysis of WEMWBS data from both minority ethnic groups in both cities, together and separately, showed a poor fit to Rasch model assumptions. This was no surprise, as WEMWBS data from the majority population in the UK also showed a poor fit with this model (Stewart-Brown et al., 2009). The latter identified a seven-item scale, called SWEMWBS (the shortened WEMWBS), that met Rasch criteria well. This shortened scale is now being used in many surveys where respondent burden is an issue.

The fit of SWEMWBS minority ethnic data to the Rasch model was much better that the fit of WEMWBS data. This was true both in individual minority groups and in the combined dataset. Rasch analysis identifies items that show differential item functioning (DIF), that is, they seem to be answered in different ways by different groups relative to their overall responses or total scores. In the original analysis (Stewart-Brown et al. 2009), for example, I found that the item "I've been feeling confident" showed significant DIF for gender (in particular, men were more likely to report more confidence). This finding seems to replicate past observations and, at some level, reassures that the scale is working well. At the mathematical level, however, it caused problems in that the item needed to be abandoned.

Construct validity of WEMWBS was tested by running a Confirmatory Factor Analysis (CFA). Statistical Analysis Software (SAS) was used for evaluating the one-factor structure. Initially no dependency between residuals was assumed, but then it was added gradually in order to obtain an acceptable model fit. In this way all the alternative indices of fit reported a good fit.

Lloyd and Devine (2012) noticed that the validation made by Tennant and colleagues in 2007 excluded the population of two constituent regions of the UK (Wales and Northern Ireland). They decided therefore to test the psychometric properties of the WEMWBS on a random sample of the general population of Northern Ireland, also to evaluate the impact that the recent Irish civil conflicts could have had on the mental health and well-being of the population.

The data came from the 2009/2010 Continuous Household Survey (CHS), which was conducted by the Central Survey Unit of the Northern Ireland Statistics and Research Agency (NISRA). The analyses were carried out on the replies of 3,355 people (aged 16 years and over) who fully completed the WEMWBS. The results showed that the overall median score was 50 (in the previous study of Tennant it was 51), and also the internal consistency was high and similar to the one reported for the English and Scottish population, confirming the

reliability of the tool. The data were analysed using an exploratory factor analysis, and in line with the research of Tennant scores showed a single underlying factor which explained 54% of the variance. The 2009/2010 CHS used a series of standardized measures, included the GHQ12, to test the criterion validity, and consistent with the results of Tennant et al., there was a statistically significant negative correlation between the two procedures. In line with Tennant, all other differences across medians in relation to age, tenure, self-perceived health status, employment status, marital status and terminal education age were statistically significant. In conclusion the finding that levels of mental well-being in Northern Ireland were similar to those in Scotland was remarkable, given the fact that at the time of the survey the two lands were experiencing different social and historical circumstances (between the others, the civil conflict in Northern Ireland and the major economic recession in UK).

Deary, Watson, Booth and Gale (2012) investigated, using Mokken scaling, if the 14 item of the WEMWBS forms a hierarchy and whether the hierarchy varies according to cognitive ability.

Cognitive ability of the subjects was assessed when the participants were 11 years using a general cognitive ability test, then part of them took part in the 2008-2009 follow-up survey, when they were 50 years, and completed the WEMWBS. The 8643 participants were divided into 2 groups according to gender and then divided into low, medium or high cognitive ability. The results of Mokken scale show a moderately strong unidimensional hierarchy of items under the model of MMH, except for the female of medium and high mental ability for which a strong hierarchy is shown. Acceptable Invariant Item Ordering (IIO) was shown for all except the low cognitive ability participants in the total sample, in the male sample and in the female one.

Items 4 and 9 were the third and the fourth most endorsed items by female but they were the fourteenth and twelfth most endorsed by males participant. Items 4,8, and 14 only show IIO in one scale each, and this is not consistent across the sub-groups of the analysis. Item 10 does not show IIO in any sub-group. Therefore, the WEMWBS does have a hierarchy of items and in all of the scales, the ordering of items is broadly similar.

Bianco (2011) examined the performance of WEMWBS for screening depression in Italian and English populations and tried to find cut-off points for identifying cases of depression and psychological distress. In order to achieve these goals, WEMWBS scores were compared with scores of the Centre for Epidemiologic Studies Depression Scale (CES-D), selected as a gold standard, through the method of ROC curves, which analyzes the relationship between true positives (or “sensitivity”) and false positives (or “specificity”). In the English sample, cut-off scores of 44.5 and 40.5 showed optimal screening properties both at the baseline and at the

two follow-ups. In the Italian sample, the cut-off point of 44.5 on the WEMWBS was able to discriminate, in a statistically significant way, between psychologically distressed and non-distressed individuals. Moreover, ANOVA showed significant differences between "positive" and "negative" subjects in the level of mental distress and depression as measured by the PGWBI and the WHO-5, while no significant influence of "gender" was found.

Ragonesi (2012) examined WEMWBS performance in screening subjects at risk of Post-Natal Depression, using the Edinburgh Postnatal Depression Scale (EPDS, Cox et al., 1987) as a gold standard. They found out that WEMWBS performed well in screening post-natal depression, by discriminating who are at risk from who are not. The most appropriated cut-off point to screen for both high and moderate likelihood of post-natal depression in new mothers seems to be 48.5 (out of the maximum score of 70). Even if the main finding was that WEMWBS may be better than EPDS in screening Post-Natal Depression (probably because for women it might be easier to admit the presence of a positive emotion), the percentage of women at risk was higher than the one indicates in the literature, possibly due to the fact that this disorder may be under diagnosed and because authors preferred to maximize the ability of WEMWBS to detect true positive cases.

### 1.3.3. The Italian validation study

In 2011, Gremigni and Stewart-Brown validated the WEMWBS for the Italian population.

Two independent groups of people participated. The first one consisted of 345 people from the North East of Italy of whom 46.7% male and 53.3% female aged between 18 and 82 years, and with an average of 14 years of education. As far as age is concerned the sample was divided in 5 groups: from 18 to 24, 25-30, 31-42 and 43-57 and over 57. Regarding the occupation, the sample was divided in four categories: unemployed, employed, retired and students. After one year the second group was recruited in a doctor's surgery in order to perform the test-retest as proof of reliability. It was composed by 52 people of whom 26 were male and 26 were female. Ages ranged from 19 to 78 and average years spent in education was 13 years.

The WEMWBS was initially translated with multiple forward translation and back translation. The scale was administered to the first group of participants together with a questionnaire about socio-demographic data and an array of questionnaires (PANAS, SWLS, PWBS, WHO-5, PGWBI) for evaluating the criterion validity. After collecting the data, the normality of the distribution of the scores were verified by checking for skewness and kurtosis between 1 and -1 as indicators of distribution that is close to normality. The WEMWBS responses were also evaluated for floor and ceiling effects using as criteria for substantial and

unacceptable effect a situation where 25% subjects chose the lowest or highest option for every item.

To verify the one factor structure of the WEMWBS a Confirmatory Factor Analysis (CFA) was run using methods and indices analogous to the English validation study (Tennant et al., 2009). Confirmatory factor analysis supported the monofactorial solution that was found in the English study although it indicates a reduction from 14 items in the original scale to 12, by eliminating item 4, corresponding to the item 8 of the English version (I've been feeling good about myself) and item 12 (I've been feeling loved). Even in the original study by Tennant and colleagues there was some redundancy of items and the authors suggested a future reduction. Reliability as internal consistency was evaluated with Cronbach's Alpha and the stability over time with test-retest after one week in an independent group of 52 participants using intra class correlation coefficient ( ICC) as in the English study.

The distribution of scores of the WEMWBS items did not show skewness or kurtosis values greater than those indicating approximation to normality (with values from -0.86 to 0.12 for skewness, and from -0.71 to 0.91 for kurtosis) As far as ceiling effects are concerned, the frequency of extreme high responses were similar to the English samples and all of the response categories were used at least one person for all the items. However, item 12 showed a frequency greater than 25% of extreme responses for the highest category

The corrected item total correlations varied from 0.28 to o.69 with a mean of r=0.51. The two items which were different from the general trend of the correlations (items 4 and 12) showed a correlation with the total of r=0.28, and a correlation with all other items of $\geq$ 0.44. Eliminating these two items the mean item-total correlation rose (r=0.54).

Analysis was conducted on two bi-factorial models, one with independent factors and the other with correlated factors. The first factor consisted of five items which formed a cluster and related to the hedonic model while the second factor consisting of seven items related to the eudaimonic model. Neither showed a good fit to the empirical data. It was considered therefore that the monofactorial model with 12 items, where saturation of the variables observed on one factor ranged from 0.39 to 0.67, was the best one. This version presented a total score that varies from 14 to 58 with a mean score of 42.06, SD = 6.59 and a distribution not too different from normal (skewness = -0.60; kurtosis=0.84) in the first group (n=345). The second group (n=52) has scores similar to the first group that vary from 20 to 56 with mean score = 42 SD=5.94 and distribution not far from normal (skewness=0.57, kurtosis=0.69). Differences between the two groups are not significant. This fact is indicative of good stability of WEMWBS when applied to different groups and at different times.

The reliability of the 12 item Italian version has a value of 0.86 for (Cronbach's) alpha. For the first group (n=345) and alpha=0.83 for the second group (n=52), values that show a good internal consistency. Stability after one week measured in the second group seems adequate with an intra-class correlation coefficient ICC = 0.70 (95% CI 0.54-0.79; F-test with value =0: F(51) = 3.9, p<0.0001). Social desirability is moderate in the Italian study similar to the English one and it has good reliability characteristics.

High correlations (r=0.26-0.62), even though lower than those of the English study, were found between WEMWBS and scales that measure hedonic wellbeing (PANAS-P, SWLS) and eudaimonic wellbeing (PWBS) and emotional wellbeing (WHO-5) and positive aspects of mental wellbeing (PGWBI: positivity and wellbeing, self-control, vitality and general health) as expected. High correlations (r=0.48-0.61) were observed in the expected direction between WEMWBS and opposite constructs such as negative affectivity (PANAS-N) and the scales of anxiety and depression of the PGWBI, while the correlation is more modest for the global health index of the GHQ12 (r= -0.17).

Educational level, measured as years of study, does not seem to be associated with the total score of WEMWBS (r= -0.04, p=0.46). Regarding the other socio-demographic variables, ANOVA model that includes gender, age and occupational status, as independent variables appear to be significant (F(19,325) = 3.22, p <0.0001, $\eta^2$ = 0.16). Interaction between gender and occupational status (F(2,325) = 2.57, p= 0.08 $\eta^2$ = 0.02), between age group and work status and of the three variables that were independent from each other were all non-significant.

To understand better the age effect differences between different age groups were calculated separately for the two genders. Females aged 25-30 had higher scores, while among males the oldest ones (>57yrs) had the highest scores. Post hoc comparisons with Bonferroni test showed that women between 25-30 years had significantly higher scores than both 18-24 yr. (mean difference 5.19, p=0.01, d=0.84) and 31-42 yr. groups (mean difference 5.25, p=0.02, d=0.79). Men aged over 57 yrs. had significantly higher scores than both males from 18-24 yrs. (mean difference=4.01, p=0.04, d=0.69) and males from 43 to 57 yrs. (mean difference=5.43, p=0.005, d=0.86).

In conclusion, its use in Italy appears to be appropriate because it presents a good internal validity, external validity, good reliability, moderate social desirability, and also criterion validity was confirmed. According to the authors, future studies should investigate the sensibility of the instrument to positive change, in order to make WEMWBS an instrument suitable for evaluating changes dues to therapeutic intervention.

### 1.3.4. A short form: the SWEMWBS

In 2009, Stewart-Brown, Tennant A., Tennant R., Platt, Parkinson, and Weich ran a Rasch analysis on data of WEMWBS obtained from the Scottish Health Education Population Survey. The Rasch model shows what should be expected in responses to items if measurement at the metric level is to be achieved (Tennant, 2007), and verifies whether the scale works in the same way regardless which group is being assessed. As the initial fit to model was poor, and because some items showed bias for gender or age, the authors decided to delete some of them (such as "I've been feeling good about myself" and "I've been feeling cheerful"). This led to a shorter version of WEMWBS, called Short Warwick-Edinburgh Mental Well-Being Scale.

The Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) is a unidimensional seven item scale that measures positive mental well-being. Its seven items were extracted by the longer version and corresponds to the item numbers 1, 2,3,6,7,9,11. It is intended for adults aged 16 years and more, and it takes less than 5 minutes to be self-completed. As in the WEMWBS, participants are asked to describes their experience over the last 2 weeks, by ticking one of the five boxes next to each sentence: 1 (None of the time), 2 (Rarely), 3 (Some of the time), 4 (Often), 5 (All of the time). Total score ranges from 7 to 35.

Most items of SWEMWBS cover aspects of eudemonic well-being (psychological functioning, positive relationship with others and self-realisation) and few covering hedonic ones. Moreover, these seven items seem to relate more to functioning than to feeling. So, SWEMWBS presents a more restricted view on mental well-being but robust measurement properties. In fact, the Rasch analysis conducted in the first study (Stewart-Brown et al., 2009) showed also scores can be turned into a metric. Metric score conversion of SWEMWBS is presented in Table 2. The correlation between the two versions of the tool was 0.954 (ibidem).

SWEMWBS has been used in different fields, included a retrospective study of the impact of childhood experiences of violence on adult well-being (Bellis et al., 2013).

***Table 2. Raw score to metric score conversion table for SWEMWBS***

| Raw score | Metric score | Raw Score | Metric Score | Raw Score | Metric Score |
|---|---|---|---|---|---|
| 7 | 7.00 | 17 | 16.88 | 27 | 24.11 |
| 8 | 9.51 | 18 | 17.43 | 28 | 25.03 |
| 9 | 11.25 | 19 | 17.98 | 29 | 26.02 |
| 10 | 12.40 | 20 | 18.59 | 30 | 27.03 |
| 11 | 13.33 | 21 | 19.25 | 31 | 28.13 |
| 12 | 14.08 | 22 | 19.98 | 32 | 29.31 |
| 13 | 14.75 | 23 | 20.73 | 33 | 30.70 |
| 14 | 15.32 | 24 | 21.54 | 34 | 32.55 |
| 15 | 15.84 | 25 | 22.35 | 35 | 25.00 |
| 16 | 16.36 | 26 | 23.21 | | |

### 1.3.5. Cross-cultural validation of the WEMWBS and the SWEMWBS

Taggart and colleagues (2013) validated the WEMWBS among two of the minority ethnic groups living in the UK, Chinese and Pakistani. The study was the first attempt to validate a mental well-being scale among minority ethnic groups. The aim was to assess the extent to which the WEMWBS is suitable for measuring mental well-being in groups in which different language, culture and beliefs may affect mental health and well-being in a different way. The researchers undertook both quantitative surveys and focus group discussions in age- and sex- specific groups. As far as the quantitative evaluation is concerned, the samples were recruited in Birmingham and Coventry in collaboration with the local Primary Care Trust (PCT) and were composed by English speaking adults living in the UK and self-identified as Chinese or Pakistani. The Birmingham sample received a booklet containing the WEMWBS, the GHQ-12 (Goldberg &Williams, 1988), the World Health Organisation Well-being 5 questionnaire (WHO-5) (Heun et al., 2001), and a demographic questionnaire. The Coventry sample completed a questionnaire suitable for a general health survey including WEMWBS, but not the GHQ-12 or the WHO-5. The data from Birmingham and Coventry samples were combined when available for both groups, even if Chinese and Pakistani scores were analysed separately. The distribution of responses showed median scores of 50 for the Chinese sample, and of 51 for the Pakistan sample, and these are similar to the median score of 51 of the Scottish population reported by Tennant et al. (2007), and to the median score of 50 in the Northern Ireland found by Lloyd and Devine (2012). Cronbach's alpha was 0.92 and 0.91 respectively for the Chinese and Pakistani samples, and is thus consistent with the previous

studies in other population samples. The item total correlation were evaluated through the Spearman rank correlation coefficients calculated for each item, and were comparable with the findings of Tennant et al. (2007). Factor analysis showed that for the Chinese sample only one factor was significant and explained 52% of the variance. For the Pakistani sample three factors were significant, explaining 48%, 8% and 7% of the variance respectively. These findings are also consistent with those in other study population.

In Birmingham sample Spearman's correlation for the WEMWBS with other scales was calculated. In the Chinese data the WEMWBS correlation with the GHQ-12 was -0.63 and with the WHO-5 it was 0.62. In the Pakistani data the correlation between WEMWBS and GHQ was -0.55 and with the WHO-5 was 0.64. In the study of Tennant and colleagues (2007) the correlation of the WEMWBS with the GHQ-12 was -0.53.

As far as the qualitative evaluation is concerned, focus groups were held in Birmingham and 22 Chinese and 47 Pakistani adults took part. Chinese participant were divided in three mixed sex groups according to age (16 to 24 years old, 25 to 49, and 50 to 75). Pakistani adults were divided by sex (in order to make women's views more likely to be heard) and by age. The topic guide was designed to identify the level of comprehension, acceptability and the clearness of the WEMWBS perceived by the participants of two minority ethnic groups, and to investigate the concepts of mental well-being relevant to each community. The discussions were recorded and transcribed. The responses relating to items were analysed item by item, while the discussion regarding mental well-being was analysed by themes. All items have been considered easy to understand, with the possible exception of item 1 among the Pakistan women (probably because there is no an exact translation for the word "optimist" in Pashto, the native language of Pashtun people of South-Central Asia). Young men in both Chinese and Pakistani groups interpreted the item "feeling interested in other people" in a sexual context. Item 2 and 3 ("I've been feeling useful" and "I've been feeling relaxed") were considered difficult to answer by some participants. Some participants observed that the tool was good to complete because it made you think about your life in terms of mental well-being.

Regarding the concept of well-being, Chinese participants in general espoused an internal model of mental health, believing that it depended on your own actions and attitudes. They also tended to be dismissive of depression believing that it was over diagnosed in England. On the other side Pakistani reported a more social model of mental health, referring to the fact that worries about the family and financial problem are the major cause of mental suffering. Moreover both men and women in Pakistani groups mentioned the spiritual interpretation of mental well-being. These findings are in line with those of Newbigging (2008). She found

that the Chinese understood the term happiness, but not mental well-being. Pakistanis equally did not understand mental well-being and talked more of peace of mind and contentment. Both groups understood and talked about the concept of feeling good about oneself, and freedom from worry came up as important for happiness.

In conclusion, the WEMWBS was well received by English-speaking members of both Pakistani and Chinese communities and showed high levels of consistency and reliability when compared with accepted criteria. Concerning the concept of well-being, cultural differences should be taken in consideration in the design of well-being scales.

López and colleagues (2012) first adapted WEMWBS into Spanish and performed a preliminary evaluation of its metric properties. 148 graduate students fully completed the first evaluation, and 52 of them also complete the 1-week retest administration. Four items were modified in order to be more conceptually and linguistically equivalent to the original, using the method of forward and back-translation. Cronbach's alpha (0.90), item-total score correlations (0.44-0.76), and test-retest Intraclass Correlation Coefficient (ICC) (0.84) were satisfactory. Moderate to high correlations ($r$= 0.45-0.70) were observed between the WEMWBS and validity scale, and the internal consistency measured by Guttman's Lambda 2, was 0.92. Exploratory factor analysis (EFA) suggested a structure with two highly correlated dimensions, even if limitations in sample size and item skewness recommended caution when interpreting the results. Confirmatory factor analysis suggested that the Spanish version was most likely one-dimensional, although it did not perfectly fit the one-factor model. Scales measuring positive aspects (like WHO-5 and PANAS-PS) had positive high correlation with the WEMWBS, consistent with a priori hypotheses, even if the Satisfaction With Life Scale (SWLS) correlation was lower than that obtained in the original validation study in UK.

After this preliminary study, the Spanish group (Castellví, Forero, Codony, Vilagut, … & Alonso, 2013) assessed the validity and reliability of WEMWBS in the general population. The questionnaire was administered, together with socioeconomic and health-related variables, to a total of 1900 participant aged over 15 only from the region of Catalonia. People could choose whether to complete it in Castellan or Catalan. No evidence of uniform or non-uniform DIF (Differential Item Functioning) was found regarding the languages. CFA fit the one-factor model adequately and showed a high internal consistency (Cronbach's alpha = 0.93). CFI, TLI (Tucker Lewis Index), and RMSEA showed adequate levels of fit. There were significant differences across subgroups in all socioeconomic and health-related variables assess, expect gender. For example a high association with being risk of mental disorder and lowest net familiar income was found. The Spanish version showed good

psychometric properties similar to the UK original scale, advising that the original and the adapted versions have cross-cultural equivalence. Interestingly, scores were higher for the population of Catalonia than that for Scotland general population (with the former presenting a highly skewed distribution of the score), suggesting the need of further research.

WEMWBS was also validated on a sample of 286 Dutch women with mild depressive symptoms (Ikink, Lamers, & Bolier, 2012), as a first step before the validation in the general population of Netherlands. No ceiling or floor effects were found and factor analysis confirmed a single factor structure with hedonistic and eudaimonic aspects. High reliability was found to correlate positively with other questionnaires evaluating similar constructs, and negatively with the depression questionnaire CES-D and the anxiety subscale HADS-A. The only limit recognized by the authors was the lack of items representing the component of social well-being.

## 1.4. Objectives

The main aim of the study is to test Measurement Invariance (MI) of the WEMWBS and SWEMWBS in order to provide evidence of the construct comparability of both tools across different groups based on country, gender, and age.

Indeed, MI examines whether an instrument tests the same constructs in different groups (Chen, 2007).

Often in research it is assumed that an instrument operates in the same way and that its theoretical structure is identical across different groups, even if these assumptions are not tested statistically (Byrne & Stewart, 2006). However, when researchers want to utilize a tool in different conditions (e.g. measurements conducted over time, or using diverse ways of administration or, as in our case, across different populations), they have to test MI in order to confirm that the tool measures the traits in the same way across all groups (Kim & Yoon, 2011).

Cross-group invariance is tested using Structural Equation Modeling (SEM), which analyses the pattern of correlations between a set of variables by combining path analysis, factor analysis, and regression analysis. The goal of assessing MI is to find put whether the same SEM model is applicable across groups. Testing the equivalence means to establish and test different models that assure that comparisons between groups, made on the same latent variables/s, are valid. When MI is established, it means that people who are identical on the construct measured by an instrument have the same probability of reporting the same scores regardless of their group membership. The establishment of MI is especially important in assessing group difference on a measure. For this reason I will compare WEMWBS and

SWEMWBS scores according to the three variables of country, gender and age, in order to investigate the role of these factors in influencing differences in the experience of well-being.

# Chapter 2. Method

## 2.1. Participants

### 2.1.1. The UK sample

English data were taken from the Health Survey for England 2011 (HSE) (NatCen Social Research, 2011).

HSE is a series of questionnaires designed to provide annual data about nation's health and monitor changes in health conditions.

HSE was first conducted in 1991, after a report of the Committee of Inquiry into the Future Development of the Public Health Function which identified a lack of capacity to monitor the health of the population. This led to the establishment of the Central Health Monitoring Unit, which commissioned an annual survey of health and nutrition (Mindell et al., 2012).

HSE collects detailed information on mental and physical health, and objective physical and biological measure, of people aged equal or more than 2 years. This pursues different aims: collecting data on a large scale estimate the prevalence of risk factors and monitor progresses in selected targets.

HSE is sponsored by the National Health Service (NHS) Information Centre for health and social care (IC). Target are all adults aged 16 or more and children aged 0-15, living in private residential accommodation in England (up to a maximum of 10 adults and 2 children).

In 2011, the main focus of HSE was on cardiovascular diseases and this is why topics concentrated on smoking, drinking and fruit and vegetable consumption. Since 2010, among the additional topics, also well-being was evaluated, through the administration of WEMWBS to adults aged 16 and over.

The HSE 2011 sample was randomly selected in 562 postcode sectors over all the year 2011, and in the end it included a total of 8610 adults aged 16 and over, and 2007 children aged 0-15. The household response rate was 66%.

HSE methods were face-to-face Computer Assisted Personal Interviewing (CAPI), self-completion and objective measurements.

Adults were asked to fill a self-completion booklet containing WEMWBS, the EQ-5D, and some questions regarding attitudes towards personal health and lifestyle.

Data of the SWEMWBS were obtained from the same source, but extracting only the 7 items composing the short version (items 1, 2,3,6,7,9,11 of WEMWBS).

Considering the huge sample size of the HSE (more than 10.000 data), I decided to use a stratified random sampling (run by SPSS), that permitted to casually select subjects from each age group. In this way the two country group were more numerically similar and also percentages in the subgroups.

Characteristics of the final English samples (for WEMWBS and SWEMWBS) are described in Table 3.

In the WEMWBS sample, 205 people were employed (58,6%), 88 unemployed (25,1%), and 57 retired (16,3%). Concerning the marital status, 159 participants were single (45,4%), 140 married (40,0%) and 51 were either separated, divorced, or widowed (14,6%). In the SWEMWBS sample, 295 people were employed (60,2%), 139 unemployed (28,4%), and 56 retired (11,4%). As far as the marital status is concerned, 223 individuals were single (45,5%), 208 married (42,4%) and 59 were either separated, divorced, or widowed (12,0%).

***Table 3. Characteristics of the sample***

| | English sample | | Italian sample | |
|---|---|---|---|---|
| | WEMWBS (N = 350) | SWEMWBS (N = 490) | WEMWBS (N = 338) | SWEMWBS (N = 481) |
| | N (%) | N (%) | N (%) | N (%) |
| Sex | | | | |
| Males | 145 (41.4) | 231 (47.1) | 158 (46.7) | 229 (47.6) |
| Females | 205 (58.6) | 259 (52.9) | 180 (53.3) | 252 (52.4) |
| Age | | | | |
| 16-34 | 170 (48.6) | 275 (56.1) | 165 (48.8) | 273 (56.8) |
| 35-54 | 90 (25.7) | 120 (24.5) | 86 (25.4) | 115 (23.9) |
| ≥ 55 | 90 (25.7) | 95 (19.4) | 87 (25.7) | 93 (19.3) |

### 2.1.2. The Italian sample

Italian data were obtained from a general population sample. They were collected by 3 psychologists, partly through acquaintances, and partly from the waiting room of a general practitioner.

Italian data of the SWEMWBS were obtained from two merged files. The first one contained the scores of the administration of SWEMWBS, and the second one was gained (similarly to

the English SWEMWBS data) by extracting from the WEMWBS dataset only the scores of the 7 items of the short version.

In the WEMWBS sample, 85 participants were students (25,1%), 194 employed (57,4%), 12 unemployed (3,6%), and 47 were retired (13,9%). Regarding the marital status, 133 subjects ere single (39,3%) and 205 married (60,7 %).

In the SWEMWBS sample, only data on marital status were available. Two hundred and thirty-two participants were single (48,2%), 248 married (51,6%), and 1 was either separated, divorced or widowed (0,2%).

Characteristics of the final Italian samples (for WEMWBS and SWEMWBS) are described in Table 3.

## 2.2. Measures

In the English sample, WEMWBS was administered together with other questionnaires, as it was part of the HSE 2011 (described above). WEMWBS (Tennant, 2007) is a scale of 14 positively worded items created in 2007 by the Universities of Edinburgh and Warwick for measuring well-being. Items cover both hedonic and eudemonic aspects, and thanks to its brevity and simplicity it was included in two National Scottish Surveys and from 2010 also in the Health Survey for England.

In 2009 , through a Rasch Analysis of WEMWBS, a short version composed by 7 items was developed. Items of both scales present a 5-point Likert Scale, so that the total score ranges from 14 to 70 in WEMWBS, and from 7 to 35 in SWEMWBS. For further information regarding psychometric properties of the measures, see Paragraph 1.3.

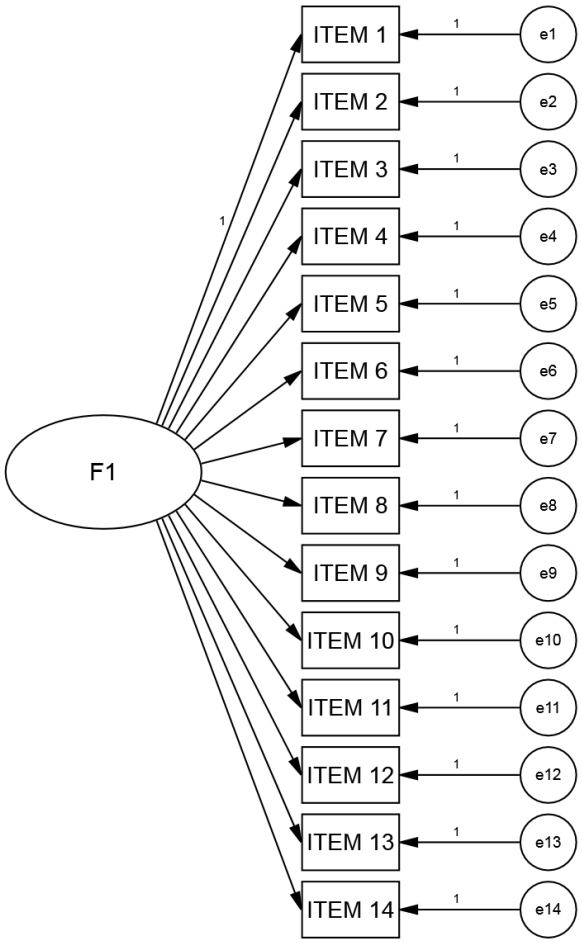
## 2.3. Statistical analysis

Before conducting MGCFAs two CFAs with the Maximum Likelihood method (ML) were performed to verify the structure of the model in UK and Italy, respectively, and to test if the baseline model (i.e., the one that holds for all groups involved and shows that the items configuration is the same) adequately represents data from each country. The ML estimation method chooses those parameters that maximize the probability of getting the observed data matrix (Barbaranelli, 2006). The path model for the hypothesized one-factor model of the WEMWBS is shown in Figure 1. The latent variable (F1) is represented by an oval whereas the 14 observed variables (i.e., the 12 items of WEMWBS are indicated by rectangles. One-way arrows draw paths from the common factor to the observed variables, as in SEM it is assumed that latent variables underlie the manifest ones, as it represents the common variance

that the 14 indicators share. Each observed variable will have a measurement error (estimated for the value 1) representing the variation of that particular item. Additional 14 unobserved variables represent the measurement errors that are specific to each item. Measurement error (also called error variance or indicator unreliability) is the unique variance of each item that is not accounted for by the latent factor (Comşa, 2010).

For SWEMWBS the same one-factor model was tested, yet with observed variables.

***Figure 1. Path diagram of WEMWBS 1***




Goodness of fit of the tested factor model was assessed using both relative and absolute goodness-of-fit indices. As Chi-square is highly sensitive to sample size it is recommended alternative fit indices (Meade et al., 2008; Barbaranelli, 2006; Cheung, 2002) to be used.

Absolute fit indices evaluate the fit of a model in reference to perfect fit, so they assess the discrepancy between the estimated and the observed covariance matrix (Jöreskog and Sörbom, 1984). The absolute fit indices considered in the present study are:

- GFI (Goodness of Fit Index) and AGFI (Adjusted Goodness of Fit Index):
  Both indices show to which extent the original matrix is explained by the reproduced matrix, but AGFI is adjusted in respect to the number of variables and degrees of freedom.

Their value goes from zero to one, where less than -9 indicate the need of extracting more factors, while value greater than .9 indicate an acceptable fit (Barbaranelli, 2006).

- SRMR (Standardized Root Mean Square Residual):

  SRMR represents the standardized difference between the observed correlation and the predicted correlation. A value of zero indicates perfect fit, and less than .08 is generally considered a good fit (Hu & Bentler, 1999). This value tends to be smaller as the sample size increases and as the number of parameters in the model increases.

Relative fit indices compare the predicted model with a null model, that a model where covariances among all input indicators are fixed to zero. The relative fit indices considered in the present study were:

- CFI (Comparative Fit Index):

  it compares the predicted model with a baseline model, which is a null or independence model in which the covariances among all input indicators are fixed to zero. In this case the presented model is compared to the worst possible model. It is a revised form of the Normed Fit Index (NFI) which takes into account the sample size and performs well even when the sample size is small. (Tabachnick, 2012). Values > 0.90 indicate a good fit (Hoe, 2008).

- RMSEA (Root Means Square Error of Approximation):

  Error of approximation is the lack of fit of the model to population data, when parameters are optimally chosen. It is more recommended than the error of estimation (lack of fit of the model when parameters are chosen via fitting to the sample data), because is less affected by sample size. According to Kline (2011) a value $\leq$ .05 shows a close approximate fit, values $\geq$.05 but < .08 indicate a reasonable approximate fit, and if $\geq$.10 they reflect a poor fit.

In addition to goodness of fit indices, modification indices were inspected to evaluate model fit. The modification index is an estimate of the amount by which the discrepancy function (between the hypothesized model and the observed data) would decrease if the constraints on that parameter were removed (Sörbom, 1989). As recommended by Brown (2006) and Jaccard & Wan (1996), modification indices of 3.84 or greater reflect the critical values of a chi-squared test with 1 degree of freedom at $p<0.05$.

Subsequent to CFAs, 3 separate MGCFAs were performed to test that the one-factor structure of WEMWBS and SWEMWBS was the same in both countries and across gender and age.

MGCFA is the most used technique for testing MI and has the advantage of accounting for measurement errors (Comşa, 2010; Byrne, 2004).

In MGCFA different models are tested, starting from the baseline model and successively using constrained nestle models, so that increasingly restrictive models are added.
In the present study, four different models were evaluated, which are called with different names according to the literature (Harrington, 2008).

- Model 1 (unconstrained model/configural invariance: it assesses if the structure of the scale (the number of factors and the items per factor) is plausible for all the groups. It is the less restrictive model because it evaluates only the extent to which the same configuration freely estimated parameters holds across groups. It implies that the measurement model hold across country/time (similar pattern) but that comparison of the measures are still not meaningful.
- Model 2 (Measurement weights/ equal factor loadings/metric invariance): it assesses if factor loadings of the items are equal across groups, that is model 2 is fixing only the factor loadings, thus evaluating to which extent the items are answered similarly (unbiased) by different groups.
- Model 3 (structural (co)variances): it tests whether factor variances and covariances of the latent variable are equal across groups. In this one-factor model variance of the latent variable was constrained to be equal across groups.
- Model 4 (measurement residuals): It can be difficult to obtain acceptable indices at the level, because this model constrains the residuals (i.e., errors) to test whether all group differences on the items are exclusively due to group differences on the common factors. It fixes error variances to be equal across groups.

Nested models were compared with a Chi-square difference test. However, since Chi-square difference test is highly sensitive to sample size (i.e., in large samples even small differences can be detected), alternative fit indices are also recommended (Meade et al., 2008; Barbaranelli, 2006; Cheung, 2002). These indices are the CFI, SRMR, RMSEA (described above), and the ΔCFI. ΔCFI shows the difference between the CFI values of two models (one with less parameters), in order to assess whether the fit of the more restrictive model is still acceptable. Its values should be less than .01 (Cheung and Rensvold, 2002).

Reliability (i.e., the extent to which a tool consistently reflects the constructs that is measuring, Field, 2009) of WEMWBS and SWEMWBS was evaluated separately for Italian and English samples. As an index of internal consistency, Cronbach's alpha coefficient was computed. Cronbach's alpha is an index of reliability associated with the variation accounted for by the true score of the underlying construct (Santos, 1999), and helps determine if it is justifiable to interpret score that have been added together. Cronbach's alpha coefficient is

acceptable if > .70 (Nunnally & Bernstein, 1978), yet a high value (i.e., if ≥ .90= might suggest that the items are overly redundant and the construct is too specific (Briggs & Cheek, 1986). For evaluating internal homogeneity, adjusted item-total correlations should have a value of at least 0.40, as recommended by Streiner & Norman (2008).

Finally, one-way analyses of variance (ANOVA) were conducted to compare WEMWBS scores between countries, genders, and age groups (three age groups: 16-34, 35-54 and over 55 years old). In case of a significant main effect of the fixed factor "age", post-hoc pairwise comparison (Bonferroni) were performed.

MGCFAs were conducted using IBM SPSS AMOS 19.0, while reliability analysis (Streiner & Norman, 2008) and one-way ANOVAs were performed using IBM SPSS Statistics 21.0.

# Chapter 3. Results

## 3.1. WEMWBS

### 3.1.1. CFAs

A CFA was performed on data from each country to test the fit of the one-factor model (Figure 1). Results of the CFA on English and Italian sample are reported in Table 4.

***Table 4. First CFA in English and Italian samples***

| Goodness of fit indices (one-factor model) | UK | Italy |
|---|---|---|
| $\chi^{2\,a}$ | 335.01 | 377.32 |
| $\chi^2$/ df | 4.35 | 4.90 |
| RMSEA (CI 90%) | .09 | .11 |
| SRMR | .05 | .07 |
| GFI | .88 | .84 |
| CFI | .88 | .80 |

[a] df = 77 ; p < 0,001

In Table 5, modification indices of the CFA with English and Italian data are shown. It is recommended to look at values higher than 3.84, starting from the highest indices. A modification index higher than 3.84 means that, if that parameter was freed, then Chi square would decrease significantly and the model fit would improve in a significant way as well.

***Table 5.Modification Indices of the CFAs with English and Italian data***

| English Sample | | | Italian Sample | | |
|---|---|---|---|---|---|
| Covariances | Modification Index | Par. Change | Covariances | Modification Index | Par. Change |
| **e8 <--> e10** | **43.49** | **.13** | **e4 <--> e9** | **66.04** | **32** |
| **e6 <--> e7** | **37.26** | **.13** | e6 <--> e11 | 25.79 | .11 |
| **e4 <--> e9** | **31.34** | **.16** | e5 <--> e13 | 23.66 | .22 |
| **e9 <--> e12** | **23.53** | **.13** | e7 <--> e11 | 20.48 | .10 |
| e3 <--> e5 | 14.51 | .13 | e3 <--> e14 | 19.64 | .11 |
| e6 <--> e9 | 12.32 | -.08 | e1 <--> e11 | 16.97 | -.12 |
| e12 <--> e14 | 11.76 | .07 | e4 <--> e10 | 15.25 | -.13 |
| e13 <--> e14 | 11.23 | .07 | e4 <--> e6 | 14.74 | -.12 |
| e3 <--> e10 | 10.64 | -.08 | e3 <--> e11 | 13.71 | -.10 |
| e4 <--> e8 | 9.84 | -.07 | e7 <--> e14 | 13.54 | -.07 |
| e11 <--> e12 | 8.52 | .07 | e10 <--> e11 | 12.69 | .08 |
| e5 <--> e10 | 8.49 | -.08 | **e6 <--> e7** | **12.26** | **.07** |
| e7 <--> e11 | 8.31 | .06 | e4 <--> e13 | 11.64 | .16 |
| e1 <--> e4 | 8.26 | .09 | e3 <--> e8 | 10.40 | .09 |
| e6 <--> e13 | 6.12 | -.06 | e11 <--> e14 | 9.42 | -.06 |
| e2 <--> e6 | 6.04 | .05 | e5 <--> e14 | 9.08 | .09 |
| e8 <--> e12 | 5.58 | -.05 | **e9 <--> e12** | **8.86** | **.10** |
| e8 <--> e9 | 5.22 | -.05 | e3 <--> e10 | 8.64 | -.08 |
| e7 <--> e13 | 5.11 | -.05 | e1 <--> e7 | 8.51 | -.08 |
| e1 <--> e11 | 4.87 | -.05 | e4 <--> e11 | 8.05 | -.10 |
| e7 <--> e10 | 4.74 | -.04 | e1 <--> e13 | 7.75 | .11 |
| e6 <--> e11 | 4.71 | .05 | e8 <--> e13 | 7.70 | -.08 |
| e8 <--> e13 | 4.27 | -.05 | e1 <--> e5 | 7.08 | .11 |
| e9 <--> e10 | 4.27 | .05 | e5 <--> e9 | 6.19 | -.097 |
| e6 <--> e12 | 4.13 | -.05 | e3 <--> e5 | 5.22 | .09 |
| | | | e2 <--> e7 | 4.98 | -.05 |
| | | | e3 <--> e12 | 4.97 | -.08 |
| | | | e5 <--> e8 | 4.77 | -.07 |
| | | | e1 <--> e14 | 4.67 | .06 |
| | | | e2 <--> e3 | 4.48 | -.06 |
| | | | e1 <--> e6 | 4.41 | -.06 |
| | | | e11 <--> e13 | 4.21 | -.06 |

In Table 5, Modification Indices shows by which extent the discrepancy between our model and the observed data would fall if the suggested correlation between the errors are freed. Parameter Change show by which extent approximately the estimate would become larger or smaller if the covariance between the two residuals was freed.

By including a path between items 8 and 10, the Chi square would decrease by 43.49 and the parameter estimate would increase from 0 to 0.13

As suggested by the highest values of the modification indices, correlations between the following residuals were added to the path model: errors 8 and 10, errors 6 and 7, errors 4 and 9, and errors 9 and 12. The single CFAs were then re-run. Results of the second CFAs were good in the English sample, as indicated by the goodness of fit indices (Table 6). However, in the Italian sample, almost all indices do not respect the recommended values. RMSEA is over 0.08, so according to Kline (2011) it does not show a reasonable fit, GFI is less than .9 so more factors should be extracted, and CFI is under the suggested value of .9 (Hoe, 2008). Only SRMR is still acceptable as it is under 0.08. (Hu & Bentler, 1999).

***Table 6. Second CFA with WEMWBS scores of English and Italian samples***

| Goodness of fit indices (one-factor model) | UK | Italy |
|---|---|---|
| $\chi^{2\,a}$ | 187.95 | 285.80 |
| $\chi^2$/ df | 2.57 | 3.91 |
| RMSEA (CI 90%) | .07 | .09 |
| SRMR | .04 | .06 |
| GFI | .93 | .87 |
| CFI | .94 | .86 |

[a] df = 73 ; p < 0,001

### 3.1.2. MGCFAs

The second CFA model (Figure 2) was then tested for MI.

Results of the MGCFA testing MI across country, are presented in Table 7. All goodness of fit indices respected the thresholds suggested by the literature.

The values of ΔCFI are good in the first three models, but over the recommended value of .01 in the difference between the third and the fourth model. This means that invariance is respected in the first three models (same configuration, same factor loading and same variance of the latent variable), but there is no invariance in the error variance, so fit indices suggest to stop at the level of structural invariance. In the first three models P values are under the recommended threshold of .001 (Hansen, Graham, Sobel, Shelton, Flay, & Johnson, 1987)

even if the metric invariance tested in the second model is quite close to the acceptable limit (.007).

***Figure 1. Path diagram of AMOS with correlated errors***

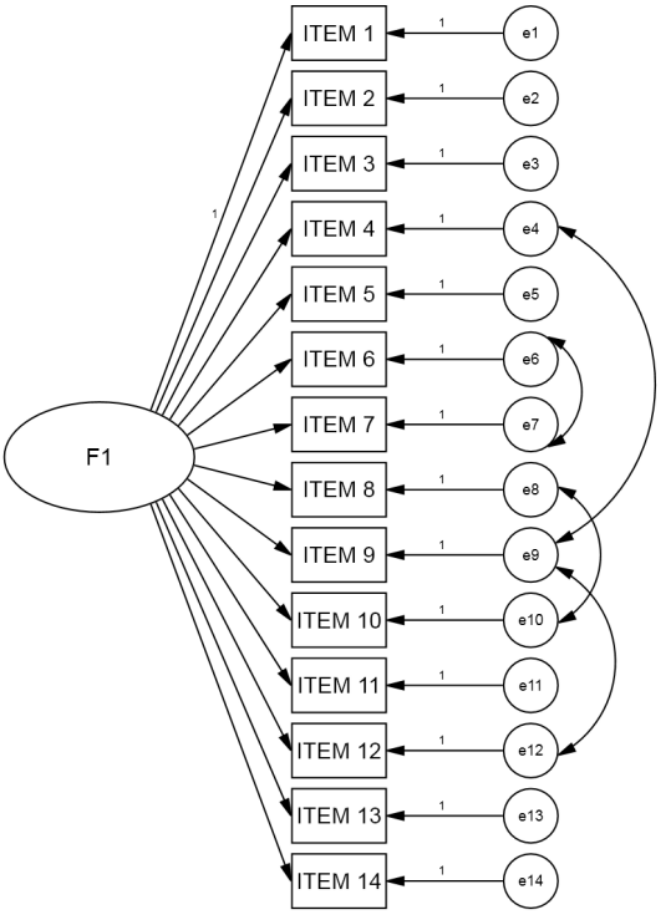



***Table 7. MGCFA of WEMWBS. MI across country.***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | p | $\chi^2$/DF | CFI | ΔCFI | RMSEA (CI 90%) | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 475.65 | - | 146 | - | - | 3.26 | .91 | - | .06 | .04 |
| Measurement weights | 504.33 | 28.68 | 159 | 13 | <.01 | 3.17 | .91 | .00 | .06 | .04 |
| Factor variances | 512.29 | 7.96 | 160 | 1 | <.001 | 3.20 | .91 | .00 | .06 | .05 |
| Measurement residuals | 601.35 | 89.06 | 178 | 18 | <.001 | 3.38 | .89 | .02 | .06 | .07 |

MGCFA across gender was performed by comparing females of both countries with males of both countries, to see whether the invariance of the tool is maintained also across gender, regardless of country. Results of the MGCFA testing MI across gender are presented in Table 8. The first three models present acceptable values of all the goodness of fit indices according

to the limits suggested by the literature, so invariance is respected till the structural invariance model. Thanks to metric invariance, it is possible to compare model parameters across group. This means invariance is respected in the first three models (same configuration, same factor loading and same variance of the latent variable), but there is no invariance in the error variance, so fit indices suggest to stop at the level of structural invariance.

***Table 8. MGCFA of WEMWBS. MI across gender***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | P | $\chi^2$/DF | CFI | ΔCFI | RMSEA (CI 90%) | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 464.55 | - | 146 | - | - | 3.18 | .91 | - | .06 | .05 |
| Measurement weights | 489.11 | 24.56 | 159 | 13 | .03 | 3.08 | .91 | .00 | .05 | .05 |
| Factor variances | 489.78 | 0.67 | 160 | 1 | 0.41 | 3.06 | .91 | .00 | .05 | .05 |
| Measurement residuals | 528.11 | 38.33 | 174 | 14 | <.001 | 2.97 | .90 | 01 | .05 | .05 |

With regard to the MGCFA across age, results are showed in Table 9. Data were divided in three age groups in order to have balanced subgroups. The first group (N=335) included people aged 16 to 34 years old. The second group (N=176) was composed by people from 35 to 54 years old. The third group (N=177) comprised participants aged 55 years old and over. Values are acceptable in the first model, showing that the model structure is maintained in the three groups, but from the second model the level of significance (p) is less than .001, showing that there is no invariance.

***Table9. MGCFA of WEMWBS. MI across age groups***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | p | $\chi^2$/DF | CFI | ΔCFI | RMSEA (CI 90%) | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 490.23 | - | 219 | - | - | 2.24 | .92 | - | .04 | .06 |
| Measurement weights | 542.92 | 52.69 | 245 | 26 | <.001 | 2.22 | .92 | .00 | .04 | .07 |
| Factor variances | 545.82 | 2.9 | 247 | 2 | <.001 | 2.21 | .91 | .01 | .04 | .08 |
| Measurement residuals | 654.93 | 109.11 | 283 | 36 | <.001 | 2.31 | .90 | .01 | .04 | .07 |

### 3.1.3. Internal consistency

Reliability coefficients for the WEMWBS are reported in Table 10. In the English sample, the value of Cronbach's alpha is high, showing a high internal consistency but also a certain redundancy between the items. All corrected item-total correlations were higher than .40 except items 4 and 12, and this is in line with results of Italian validation study of WEMWBS (Gremigni & Stewart-Brown, 2011).

***Table 10. International consistency coefficients of WEMWBS in Italian and English samples***

| | UK | Italy |
|---|---|---|
| N of items | 14 | 14 |
| Cronbach's alpha | .91 | .85 |
| Item 1 | ,53 | ,52 |
| Item 2 | ,62 | ,51 |
| Item 3 | ,62 | ,44 |
| Item 4 | ,50 | ,28 |
| Item 5 | ,54 | ,49 |
| Item 6 | ,61 | ,57 |
| Item 7 | ,70 | ,54 |
| Item 8 | ,78 | ,68 |
| Item 9 | ,60 | ,49 |
| Item 10 | ,70 | ,68 |
| Item 11 | ,60 | ,48 |
| Item 12 | ,49 | ,28 |
| Item 13 | ,60 | ,48 |
| Item 14 | ,80 | ,65 |

### 3.1.4. Differences across groups

WEMWBS total scores were compared between countries, gender, and age groups by running one-way ANOVA. Table 11 reports the result of the analysis. There are no significant interactions between the three fixed factors. Significant differences were found between countries ($F(1,676) = 5,04$), $p = 0.02$) and genders ($F(1,676)$, $p = 0.02$), with Italian people ($M= 49.70$, $SD= 7.29$) scoring lower than English ($M=51.28$, $SD=8.44$), and females

(M=49.90, SD=7.97) reporting a minor level of well-being comparing to males (M=51.27, SD=7.82). In Table 12 mean scores of each group are presented.

***Table 11. Test of Between-Subjects Effects***

| Source | DF | Mean Square | F | Sig. |
|---|---|---|---|---|
| Age_Group | 2 | 40,64 | ,66 | ,52 |
| Sex | 1 | 363,49 | 5,86 | ,02 |
| Country | 1 | 312,10 | 5,04 | ,02 |
| Age Group * Sex | 2 | 5,90 | ,09 | ,91 |
| Age Group * Country | 2 | 57,85 | ,93 | ,39 |
| Sex * Country | 1 | 191,89 | 3,10 | ,08 |
| Age Group * Sex * Country | 2 | 70,07 | 1,13 | ,32 |
| Error | 676 | 61,97 | | |
| Total | 688 | | | |
| Corrected total | 687 | | | |

***Table 12. Comparison of WEMWBS scores***

| | | Mean | Std. Deviation |
|---|---|---|---|
| Country | Italy | 49.70 | 7.29 |
| | United Kingdom | 51.28 | 8.44 |
| Gender | Male | 51.27 | 7.82 |
| | Female | 49.90 | 7.97 |
| Age group | 16 → 34 | 50,56 | 7,50 |
| | 35 → 54 | 49,83 | 8,05 |
| | 55 + | 51,05 | 8,58 |

## 3.2. SWEMWBS

### 3.2.1. CFAs

A CFA was performed on SWEMWBS data from each country to test the fit of the one-factor model (Figure 1). Result of the CFA on English and Italian samples are reported in Table 13. Alternative fit indices are good according to the values recommended by the literature. Note that all analyses with SWEMWBS were run with the metric value of the score, i.e., scores were converted (as suggested by the authors) according to the table reported in Chapter 1.

***Table 4. First CFA with SWEMWBS scores of English and Italian samples***

| Goodness of fit indices (one-factor model) | UK | Italy |
|---|---|---|
| $\chi^{2\,a}$ | 65.63 | 70.72 |
| $\chi^2$/ df | 4.69 | 5.05 |
| RMSEA (CI 90%) | .09 | .09 |
| SRMR | .04 | .05 |
| GFI | .96 | 1.00 |
| CFI | .96 | .93 |

[a] df=14 p<.001

In Table 14, modification indices of the CFA with English and Italian data are shown. It is recommended to look at values higher than 3.84, starting from the highest indices. A modification index higher than 3.84 means that if that parameter was freed then Chi square would decrease significantly and the model fit would improve in a significant way as well.
By including a path between items 1 and 2, the Chi square would decrease by 30.81 and the parameter estimate would increase from 0 to 0.12.

***Table 14. CFAs with SWEMWBS score of English and Italian samples***

| English Sample | | | Italian Sample | | |
|---|---|---|---|---|---|
| Covariances | Modification Index | Par. Change | Covariances | Modification Index | Par. Change |
| e1 <--> e2 | 30.81 | .12 | e1 <--> e2 | 14.87 | .11 |
| e5 <--> e7 | 19.35 | .08 | e1 <--> e3 | 13.57 | .13 |
| e2 <--> e4 | 6.28 | -.05 | e4 <--> e6 | 12.75 | -.08 |
| e2 <--> e5 | 5.57 | -.044 | e3 <--> e7 | 12.13 | -.09 |
| e1 <--> e7 | 4.47 | -.05 | e2 <--> e5 | 9.57 | -.07 |
| e4 <--> e5 | 4.42 | .04 | e1 <--> e7 | 6.17 | -.07 |
| | | | e1 <--> e5 | 4.35 | -.05 |

As the alternative fit indices are already good, I decided to run the MGCFA across country, gender, and age without freeing any parameter concerning covariance between errors.

### 3.2.2. MGCFAs

Results of the MGCFA testing MI across country are presented in Table 15. All goodness of fit indices respected the thresholds suggested by the literature, and ΔCFI is not over .01. However, Chi square difference test is significant in the last model, showing that the error invariance is not respected.

***Table 15. MGCFA across country for SWEMWBS***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | p | $\chi^2$/DF | CFI | ΔCFI | RMSEA (CI 90%) | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 136.35 | - | 28 | - | - | 4.87 | .95 | - | .06 | .04 |
| Measurement weights | 147.82 | 11.47 | 34 | 6 | .08 | 4.35 | .94 | .01 | .06 | .04 |
| Factor variances | 149.63 | 1.81 | 35 | 1 | .18 | 4.27 | .94 | .00 | .06 | .05 |
| Measurement residuals | 187.48 | 37.85 | 42 | 7 | <.001 | 4.46 | .93 | .01 | .06 | .06 |

MGCFA across gender was performed by comparing females of both countries with males of both countries, to see whether the invariance of the tool is maintained also across gender, regardless of country. Results of the MGCFA testing MI across gender are presented in Table 16. All four models present acceptable values of all the goodness of fit indices according to the limits suggested by the literature and ΔCFI and $\Delta\chi^2$ suggest to retain each progressively stringent model. This demonstrates that SWEMWBS model is invariant across the two genders.

***Table 16. MGCFA across gender for SWEMWBS***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | p | $\chi^2$/DF | CFI | ΔCFI | RMSEA | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 111.17 | - | 28 | - | - | 3.97 | .96 | - | .05 | .03 |
| Measurement weights | 120.14 | 8.97 | 34 | 6 | .18 | 3.53 | .96 | .00 | .05 | .04 |
| Factor variances | 120.58 | 0.44 | 35 | 1 | .51 | 3.44 | .96 | .00 | .05 | .04 |
| Measurement residuals | 135.89 | 15.31 | 42 | 7 | .03 | 3.23 | .95 | .01 | .05 | .04 |

With regard to the MGCFA across age, results are showed in Table 17. Data were divided in three age groups: the first (N=548), included people from 16 to 34 years old, the second age group (N=235) consisted of participants from 35 to 54 years old, and the third group (N=188) was composed by people aged 55 and over. Values of the indices are acceptable. The ΔCFI exceed the recommended value of .01 only in the difference between the fourth and the third model. The *p* values of the Chi square difference test are non significant (showing that the two compared models are similar) till the third level. The fourth model is significantly different from the previous one, showing therefore that error invariance is not met.

***Table 17. MGCFA across age groups for SWEMWBS***

| | $\chi^2$ | $\Delta\chi^2$ | DF | ΔDF | p | $\chi^2$/DF | CFI | ΔCFI | RMSEA (CI 90%) | SRMR |
|---|---|---|---|---|---|---|---|---|---|---|
| Unconstrained | 143.40 | - | 42 | - | - | 3.41 | .95 | - | .05 | .06 |
| Measurement weights | 163.30 | 19.90 | 54 | 12 | .07 | 3.02 | .95 | .00 | .05 | .07 |
| Factor variances | 166.88 | 3.58 | 56 | 2 | .17 | 2.98 | .95 | .00 | .04 | .07 |
| Measurement residuals | 209.05 | 42.17 | 70 | 4 | <.001 | 2.99 | .93 | .02 | .04 | .08 |

### 3.2.3. Internal consistency

Reliability coefficients for the WEMWBS are reported in Table 18. The value of Cronbach's alpha is high, showing a high internal consistency. None of the corrected item-total correlations was under the recommended value of .40, suggesting adequate contribution of each item to the scale homogeneity.

***Table 18. Internal consistency coefficients of SWEMWBS in Italian and English sample***

| | UK | Italy |
|---|---|---|
| N of items | 7 | 7 |
| Cronbach's alpha | ,85 | ,80 |
| Item 1 | ,58 | ,52 |
| Item 2 | ,64 | ,52 |
| Item 3 | ,57 | ,42 |
| Item 4 | ,67 | ,66 |
| Item 5 | ,67 | ,60 |
| Item 6 | ,57 | ,46 |
| Item 7 | ,54 | ,55 |

### 3.2.4. Differences across groups

SWEMWBS total scores were compared between countries, gender, and age groups by running a single one-way ANOVA. Table 19 reports the result of the analysis. There are no significant interactions between the three fixed factors. Significant differences were found only between countries (F(1,959)=10.73, p= .001) with Italian people (M=22.67, SD=3.55) scoring less than English (M=23.54, SD=4.00). In Table 20 mean scores of each group are presented.

***Table 19. Test of Between-Subjects Effects***

| Source | DF | Mean Square | F | Sig. |
|---|---|---|---|---|
| Sex | 1 | 7,95 | ,55 | ,46 |
| Age_Group | 2 | 11,74 | ,82 | ,44 |
| Country | 1 | 153,82 | 10,73 | ,001 |
| Sex * Age_Group | 2 | 5,76 | ,40 | ,67 |
| Sex * Country | 1 | 7,79 | ,54 | ,46 |
| Age_Group * Country | 2 | 9,19 | ,64 | ,53 |
| Sex * Age_Group * Country | 2 | 16,89 | 1,18 | ,31 |
| Error | 959 | 14,33 | | |
| Total | 971 | | | |
| Corrected total | 970 | | | |

***Table 20. Means of SWEMWBS scores across country, gender and age groups***

| | | Mean | Std. Deviation |
|---|---|---|---|
| Country | Italy | 22,67 | 3,55 |
| | United Kingdom | 23,54 | 4,00 |
| Gender | Male | 23,26 | 3,91 |
| | Female | 22,97 | 3,71 |
| Age group | 16 → 34 | 22,96 | 3,89 |
| | 35 → 54 | 23,28 | 3,66 |
| | 55 + | 23,33 | 3,72 |

# Chapter 4. Discussion and conclusion

## 4.1. Discussion

The aim of the study was to test the Measurement Invariance of WEMWBS and SWEMWBS across two countries, gender, and three age groups. Results showed that:

- The one-factor structure of WEMWBS model tested as in the original study of Tennant (2007) does not show an acceptable values in all the fits indices. Fit of the model to the observed data improves after creating covariances between some of the items errors as suggested by the highest modification indices of the confirmatory factor analysis.
- SWEMWBS reported a better fit to the observed data than the WEMWBS. This was showed by the fact that the Alternative Fit Indices (AFIs, such as CFI) had an acceptable value without setting any path between residuals.
- Reliability measured as internal consistency was good in both scales, with no item-total correlation under the recommended value of 0.40. However, the high value of Cronbach's alpha in WEMWBS suggested some redundancy in the scales.
- Comparison of the mean scores of both tools indicated significant differences across countries (with a minor score in the Italian sample) and only WEMWBS showed significant differences also in gender, as males scored more than females.

As far as WEMWBS models are concerned, I had to add some paths as suggested by the modification indices. This brought to an improvement of the model fit to observed data. When many covariances need to be drawn, it may mean that items measure more than one factor (Comşa, 2010). In this case, some authors suggest to run an Exploratory Factor Analysis (Jöreskog & Söborn, 1979; Muthén & Muthén, 1998). However, in our case, the number of covariances suggested was not so large in comparison to the total number of items.

Theoretically, it made sense to modify the initial model by adding 4 covariances between the measurement errors. In fact, the two items with the highest modification index were item 8 ("I've been feeling good about myself") and item 10 ("I've been feeling confident"). These two items can be considered conceptually related as they both concern image that the person has about himself/herself. At the same way, items 6 ("I've been dealing with problems well") and 7 ("I've been thinking clearly") can be connected by the fact that a clear thinking normally facilitate the resolution of problems (indeed reasoning and problem-solving are two linked concepts). Modification indices showed that item 4 ("I've been feeling interested in

other people") had a high covariance with item 9 ("I've been feeling close to other people"), which in turn was connected with item 12 ("I've been feeling loved"), and actually all three refer to a more social and interpersonal dimension of well-being. Therefore, conceptually, it had sense to modify the initial model by including correlation between these measurement errors.

This result is consistent both with the original study by Tennant and colleagues (2007), in which the high value of Cronbach's alpha suggested some item redundancy in the scale, and with the Italian validation (Gremigni & Stewart-Brown, 2011) in which it was suggested a reduction from 14 to 12 items (and items which were deleted were items 8 and 12). In fact, a high value of Cronbach's alpha indicates that some of the items are measuring the same or very similar concepts. In this case it is more advisable to reduce the number of items, in order to have a shorter instrument (that is generally easier to administrate).

The high value of the modification indices is also consistent with the scores reported in the above-mentioned items. In fact, 63,4% of English people gave exactly the same response to both items 8 and 10 of WEMWBS (that was the pair of items with the highest modification index in the single CFA), while 57,1% of Italian people and 53,4% of English people ticked the same box in both item 4 and item 9, which was the pair of items with highest modification index presented in the single CFAs of both samples. That is where the highest redundancy of the long version of the tool is shown.

Item redundancy occurs when two or more items are investigating the same or very similar constructs. Redundancy can increase the amount of measurement error in the data (Baird & Lucas, 2011).

Also in the original validation study of Tennant and colleagues (2007), in which CFA was run using SAS software, initially no dependency between residuals was assumed, but then a matrix element representing the highest connection was added in order to obtain the best model fit.

MGCFAs showed that not all levels of invariance were satisfied. Regarding WEMWBS invariance across country was confirmed till the second model (metric invariance), invariance across gender was respected in the first three models, and while in MGCFA across age only the same factorial structure was found. As far as SWEMWBS is concerned, all three MGCFAs exhibited configural, metric and factor invariances, showing a better performance of the tool across different groups.

In general, I can thus say that in both WEMWBS and SWEMWBS the one-factor structure was confirmed, taking into account that the quite high level of redundancy of WEMWBS and a certain level of unexplained variance make the short version more advisable to use.

The fact that the test of MI of SWEMWBS reported a better fit than the one of the long version is consistent with previous research (Bartram, Sinclair, & Baldwin, 2013) showing better psychometric properties of this version. A recommendation for further research may be to privilege the use of the short version, which, beside needing less time to be complete, showed better psychometric properties. After proving the MI of both tools, I could proceed with the evaluation of the scores.

Gender- and age- related differences were assessed by keeping the two samples together (so that all females of the two countries were compared with all males). This was done because, after establishing the invariance of both tool, I was more interested in estimating dissimilarities between groups and not within. Furthermore, I preferred to keep the same group division of the previous analysis (i.e., the MGCFA).

As I could expect by the previous findings of the literature (Diener, Kahneman & Helliwell, 2010; Organisation for Economic Co-operation and Development, 2013), Italian people reported a significant lower score.

Differences in SWEMWBS scores are in line with the literature (Stewart-Brown, 2013) which displayed a lower sum in Italian sample.

To explain why there is a gender difference only in WEMWBS data set and not in SWEMWBS, it may help to have a look on WEMWBS items that are not present in the short version. It is possible then that main differences between the two genders rely on a social dimension (as shown by items 4 “I’ve been feeling interested in other people” and item 12 “I’ve been feeling loved), or in the perception the individual has about himself (as shown by items 8 “ I’ve been feeling good about myself” , item 10 “I’ve been feeling confident and also item 14 “ I’ve been feeling cheerful”), and in the attitude towards new activities (item 13 “I’ve been interested in new things” and item 5 “ I’ve had energy to spare”).

Note also that three of the four item pairs that were creating some problems in checking the MI (8 and 10,4 and 9, 9 and 12) are also the same that are presented only in the long version and not in the short one (the only item pair presented in both tools is the pair 6 and 7).

## 4.2. Limitations

Limitations of the study were:

1) The high percentage of participants in the first age group.

Almost half of the sample of the Italian data set (48.8%) was aged 16-34 years, probably due to the fact that data were collected in a University. In order to make the comparison between subgroups more balanced I selected a similar proportion for the English sample, but in this way the total sample was not equally divided.

2) The lack of multivariate normality.

The parameter estimation model which has been used was Maximum Likelihood. This technique requires the multivariate normal distribution of the observed and unobserved variables. However, Mardia's test was significant (i.e., this criterion was not met) but I decided to carry on with the analyses. An alternative could be to use the EQS statistical software (Bentler, 1985) to perform a robust-ML test that is more appropriate give no normality distributed data (Savalei & Bentler, 2010).

3) The lack of homogeneity of variances.

Variables should have the same finite, or limited, variance, in order to compare them across different groups. To test this requirement Levene's test was used for each independent variable (fixed factor). This test was non significant for age and gender, but significant for the variable country, so homogeneity was violated.

## 4.3. Conclusion

Goals of the study were to evaluate the Measurement Invariance of WEMWBS and SWEMWBS, to check the level of reliability as internal consistency of the scales through the value of the Cronbach's alpha, and to compare means in order to see differences in reported well-being according to the three dimension of country, gender and age.

I managed to test the Measurement Invariance of an instrument which evaluates Well-Being.

Further research which implies the comparison between different populations should always check this important psychometric property (the Measurement Invariance) because it confirms the soundness of cross-cultural estimations. This is especially relevant when an instrument is validated in many countries (that is the case of WEMWBS and SWEMWBS) and it has an interest in evaluating a concept on many samples. It would be therefore interesting to test this characteristic also in the other populations in which WEMWBS and SWEMWBS are used. Moreover, as the results showed, it is not a huge number of items that reinforces the validity of a tool, but its psychometric properties that indicate its usefulness and efficiency.

## References


Abbott, R. A., Croudace, T. J., Ploubidis, G. B., Kuh, D., Richards, M., & Huppert, F. A. (2008). The relationship between early personality and midlife psychological well-being: evidence from a UK birth cohort study. *Social Psychiatry and Psychiatric Epidemiology*, *43*(9), 679-687.

Acheson, D. (1988). *Public Health in England: The Report of the Committee of Inquiry into the Future Development of the Public Health Function*. HMSO, London.

Albanesi, S., & Olivetti, C. (2009). Home production, market production and the gender wage gap: Incentives and expectations. *Review of Economic Dynamics*, *12*(1), 80-107.

Andrews, F. M., & Withey, S.B. (1976). *Social Indicators of Well- being: America's Perception of Life Quality.* Plenum Press, New York.

Arbuckle, J. L. (2012). *IBM SPSS Amos 21 user's guide*. IBM, Armonk, NY.

Aspinwall, L.G. (1998). Rethinking the role of positive affect in self-regulation. *Motivation and Emotion*, *22*(1), 1-32.

Baird, B. M., & Lucas, R. E. (2011). "… And How About Now?": Effects of Item Redundancy on Contextualized Self-Reports of Personality. *Journal of Personality*, *79*(5), 1081-1112.

Barbaranelli, C., & D'Olimpio, F. (2006). *Analisi dei dati con SPSS* (Vol. 2). Milano: LED.

Bartram, D. J., Sinclair, J. M., & Baldwin, D. S. (2013). Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis. *Quality of Life Research*, 1-13.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186-203.

Bellis, M. A., Hughes, K., Jones, A., Perkins, C., & McHale, P. (2013). Childhood happiness and violence: a retrospective study of their impacts on adult well-being. *BMJ open*, *3*(9), e003427.

Bengtsson-Tops, A., & Hansson, L. (2001). The validity of Antonovsky's sense of coherence measure in a sample of schizophrenic patients living in the community. *Journal of Advanced Nursing*, *33*(4), 432-438.

Bentler, P. M. (1985). *Theory and implementation of EQS: A structural educational program*. BMDP Statistical Software ,Los Angeles, CA.

Berger, M.L., Howell, R., Nicholson, S., & Shards, C. (2003). Investing in healthy human capital. *Journal of Occupational and Environmental Medicine*, *45*(12), 1213-1225.

Bianco D. (2011). *The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development of clinical cut-off scores.* Thesis for the Master in Clinical Psychology. University of Bologna.

Bloom, D. E., & Canning, D. (2000). The health and wealth of nations. *Science (Washington), 287*(5456), 1207-1209.

Blunch, N. (2012). *Introduction to Structural Equation Modeling Using IBM SPSS Statistics and AMOS*. Sage Pub., Newbury Park, CA.

Borsa, J. C., Damásio, B. F., Bandeira, D. R., & Gremigni, P. (2013). The Peer Aggressive and Reactive Behavior Questionnaire (Parb-Q): Measurement Invariance Across Italian and Brazilian Children, Gender and Age. *Child Psychiatry & Human Development*, 1-11.

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality*, *54*(1), 106-148.

Brooks, R., Rabin, R., & De Charro, F. (2003). *The measurement and Valuation of Health Status Using EQ-5D: A European Perspective: Evidence from the EuroQoL BIO MED Research Programme.* Springer, Dordrecht, The Netherlands.

Brown, G. D., Gardner, J., Oswald, A. J., & Qian, J. (2008). Does Wage Rank Affect Employees' Well-being?. *Industrial Relations: A Journal of Economy and Society*, *47*(3), 355-389.

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

Burton, C. M., & King, L.A. (2004). The health benefits of writing about intensely positive experiences. *Journal of Research in Personality*, *38*(2), 150-163.

Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling*, *11*(2), 272-300.

Byrne, B. M. (2009). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Taylor & Francis/Routledge, New York.

Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling*, *13*(2), 287-321.

Carstens, J. A., & Spangenberg, J. J. (1997). Major depression: A breakdown in sense of coherence?. *Psychological Reports*, *80*(3c), 1211-1220.

Castellví, P., Forero, C. G., Codony, M., Vilagut, G., Brugulat, P., Medina, A., ... & Alonso, J. (2013). The Spanish version of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) is valid for use in the general population. *Quality of Life Research*, 1-12.

Cheavens, J. S., Feldman, D. B., Gum, A., Michael, S. T., & Snyder, C.R. (2006). Hope therapy in a community sample: A pilot investigation. *Social Indicators Research, 77*(1), 61-78.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Chida, Y., & Steptoe, A. (2008). Positive psychological well-being and mortality: a quantitative review of prospective observational studies. *Psychosomatic Medicine*, *70*(7), 741-756.

Clark, Andrew, Paul Frijters and Mike Shields (2008). Relative Income, Happiness and Utility: An Explanation for the Easterlin Paradox and Other Puzzles. *Journal of Economic Literature* 46(1), 95-144.

Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., ... & Stewart-Brown, S. (2011). Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health*, *11*(1), 487-487.

Comşa, M. (2010). How to Compare Means of Latent Variables across Countries and Waves: Testing for Invariance Measurement. An Application Using Eastern European Societies. *Sociologia, 42*(6), 639-669.

Coolican, H. (2009). *Research methods and statistics in psychology*. Routledge.

Cox, J.L., Holden, J.M. and Sagovsky, R. (1987). Detection of postnatal depression: development of the 10-item Edimburgh Postnatal Depression Scale. *British Journal of Psychiatry*, *150*,782-786.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Cummins, R. A. (2003). Normative life satisfaction: Measurement issues and a homeostatic model. *Social Indicators Research, 64*(2), 225-256.

Danner, D. D., Snowdon, D. A., & Friesen, W. V. (2001). Positive emotions in early life and longevity: findings from the nun study. *Journal of Personality and Social Psychology*, *80*(5), pp. 804-813.

Davidson, R. J., Kabat-Zinn, J., Schumacher, J., Rosenkranz, M., Muller, D., Santorelli, S. F., … & Scheridan, J. F. (2003). Alterations in brain and immune function produced by mindfulness meditation. *Psychosomatic Medicine, 65*(4), 564-570.

Davidson, S., Myant, K., & O'Connor, R. (2007). *Well? What do you think?(2006): The third national Scottish survey of public attitudes to mental health, mental wellbeing and mental health problems.* Scottish Executive Social Research, Edinburgh.

Deary, I. J., Watson, R., Booth, T., & Gale, C. R., (2012). Does Cognitive Ability Influence Responses to the Warwick-Edinburgh Mental Well-Being Scale?. *Psychological Assessment*. Retrieved from http://psycnet.apa.org.

Della, L. J., DeJoy, D. M., Goetzel, R. Z., Ozminkowski, R. J., & Wilson, M. G. (2008). Assessing management support for worksite health promotion: psychometric analysis of the leading by example (LBE) instrument. *American Journal of Health Promotion, 22*(5), 359-367.

Department of Health. (2011). *No health without mental health: A cross-government mental health outcomes strategy for people of all ages.* Retrieved September 30, 2013, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213761/dh_124058.pdf

Department of Health. (2012). *Healthy Lives, Healthy People: Improving Outcomes and Supporting transparency.* Retrieved October 1, 2013, from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/216160/Improving-outcomes-and-supporting-transparency-part-1A.pdf

Diener, E. (1984). Subjective well-being. *Psychological Bulletin, 95*, 542-575.

Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research, 31*(2), 103-157.

Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment, 49*(1), 71-75.

Diener, E., & Chan, M. Y. (2011). Happy People Live Longer: Subjective Well-Being Contributes to Health and Longevity. *Applied Psychology: Health and Well-Being, 3*(1), 1-43.

Diener, E., & Suh, E. M. (Eds.). (2000). *Culture and subjective well-being*. The MIT press.

Diener, E., Diener, M., & Diener, C. (1995). Factors predicting the subjective well-being of nations. *Journal of Personality and Social Psychology, 69*(5), 851-864.

Diener, E., Kahneman, D., & Helliwell, J. (2010). *International differences in well-being*. Oxford University Press, Oxford.

Diener, E., Suh, E.M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin, 125*(2), 276-302.

European Pact for Mental health and Wellbeing. (2008). Retrieved September 30, 2013, from http://ec.europa.eu/health/ph_determinants/life_style/mental/docs/pact_en.pdf

Fava, G. A., Rafanelli, C., Cazzaro, M., Conti, S., & Grandi, S. (1998). Well-being therapy. A novel psychotherapeutic approach for residual symptoms of affective disorders. *Psychological Medicine, 28*(02), 475-480.

Ferrara, A. R., & Nisticò, R. (2013). Well-Being Indicators and Convergence Across Italian Regions. *Applied Research in Quality of Life, 8*(1), 15-44.

Ferring, D., Balducci, C., Burholt, V., Wenger, C., Thissen, F., Weber, G., & Hallberg, I. (2004). Life satisfaction of older people in six European countries: findings from the European study on adult well-being. *European Journal of Ageing*, *1*(1), 15-25.

Field, A. (2009). *Discovering statistics using SPSS*. Sage Pub., Newbury Park, CA.

Fordyce, M. W. (1977). Development of a program to increase personal happiness. *Journal of Counseling Psychology, 24*(6), 511-521.

Friedli, L. (2009). *Mental health, resilience and inequalities*. Retrieved October, 2, 2013, from http://wellbeingsoutheast.org.uk/uploads/FB602619-E582-778A-9979F5BB826FDE9B/Mental%20Health,%20Resilience%20and%20Inequalities.pdf

George, D. (2003). *SPSS for Windows Step by Step: A Simple Study Guide and Reference, 17.0 Update, 10/e*. Allyn & Bacon, Boston.

Gerstorf, D., Ram, N., Mayraz, G., Hidajat, M., Lindenberger, U., Wagner, G. G., & Schupp, J. (2010). Late-life decline in well-being across adulthood in Germany, the United Kingdom, and the United States: Something is seriously wrong at the end of life. *Psychology and Aging*, *25*(2), 477-485.

Giannopoulos, V. L., & Vella-Brodrick, D. A. (2011). Effects of positive interventions and orientations to happiness on subjective well-being. *The Journal of Positive Psychology, 6*(2), 95-105.

Goldberg, D. P., & Williams, P. (1988). *A user's guide to the General Health Questionnaire.* Nfer-Nelson, Windsor, UK.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care, 44*(11 Suppl 3), S78.

Gremigni, P., & Stewart-Brown, S. (2011). Measuring mental well-being: Italian validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS).*Giornale Italiano di Psicologia, 38*(2), 485-508.

Griffiths, C. A. (2009). Sense of coherence and mental health rehabilitation. *Clinical Rehabilitation*, *23*(1), 72-78.

Hansen, W. B., Graham, J. W., Sobel, J. L., Shelton, D. R., Flay, B. R., & Johnson, C. A. (1987). The consistency of peer and parent influences on tobacco, alcohol, and

marijuana use among young adolescents. *Journal of behavioral medicine*, *10*(6), 559-579.

Hansson, K., Olsson, M., & Cederblad, M. (2004). A salutogenic investigation and treatment of conduct disorder (CD): Results from a long-term follow-up of youths diagnosed with CD investigated and treated in inpatient psychiatric care. *Nordic Journal of Psychiatry*.

Harrington, D. (2008). *Confirmatory factor analysis*. Oxford University Press., Oxford.

Heun, R., Bonsignore, M., Barkow, K., & Jessen, F. (2001). Validity of the five item WHO Well-Being Index (WHO-5) in an elderly population. *European Archives of Psychiatry and Clinical Neuroscience, 251*(2), 27-31.

HM Government (2011). *No Health Without Mental Health: a cross-government mental health outcomes strategy for people of all ages.* Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213761/dh_124058.pdf

Hoe, S. L. (2008). Issues and procedures in adopting structural equation modeling technique. *Journal of Applied Quantitative Methods*, *3*(1), 76-83.

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: a meta-analytic review. *PLoS Medicine, 7*(7), e1000316.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Eletrinic Journal of Business Research Methods, 6*(1), 53-60.

Howell, R. T., Kern, M. L., & Lyubomirsky, S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes. *Health Psychology Review*, *1*(1), 83-136.

Hoyle, R. H. (2000). Confirmatory factor analysis. In H. E. A. Tinsely & S. D. Brown (Eds.), *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (pp.465-497). Academic Press, New York.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1-55.

Huppert, F. A. , Baylis, N., & Keverne, B. (2005). *The science of well-being.* Oxford University Press, Oxford.

Huppert, F.A., & Johnson, D. M. (2010). A controlled trial of mindfulness training in schools: The importance of practice for an impact on well-being. *The Journal of Positive Psychology, 5*(4), 264-274.

Ikink, J. G., Lamers, S. M., & Bolier, J. M. (2012). De Warwick-Edinburgh Mental Well-being Scale (WEMWBS) als meetinstrument voor mentaal welbevinden in Nederland.Retrieved October 2013 from www.essay.utwente.nl.

Jaccard, J., & Wan, C. K. (Eds.). (1996). *LISREL approaches to interaction effects in multiple regression*. *Sage University Series Quantitative Applications in the Social Sciences, 114.* Sage Pub., Newbury Park, CA.

Jahoda, M. (1958). Current concepts of positive mental health. *Journal of Occupational and Environmental Medicine, 1*(10). Basic Books, New York.

Jha, A. P., Stanley, E. A., Kiyonaga, A., Wong, L., & Gelfand, L. (2010). Examining the protective effects of mindfulness training on working memory capacity and affective experience. *Emotion, 10*(1), 54-64.

John, O. P., & Benet-Martinez, V. (2000). Measurement: Reliability, construct validation, and scale construction. In H.T. Reis & C.M. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 339–369). Cambridge University Press, New York.

Joreskog, K. G., Sorbom, D., Magidson, J., Aranas-Dowling, E., Stetsenko, S. G., Pedersen, P. O., ... & Hemmasi, M. (1979). Advances in factor analysis and structural equation models. *Demografichni Doslidzhennya*, *4*(2), 78-89.

Jöreskog, K., & Sörbom, D. (1984). *LISREL VI user's guide.* (3rd ed.). Scientific Software, Mooresville, IN.

Joseph, S., Linley, P.A., Harwood, J., Lewis, C.A., & McCollam, P., (2004). Rapid assessment of well-being: The Short Depression-Happiness Scale (SDHS). *Psychology and Psychotherapy: Theory, Research and Practice, 77*(4), 463-478.

Kammann, R., & Flett, R. (1983). Affectometer 2: A scale to measure current level of general happiness. *Australian Journal of Psychology,35*(2), 259-265.

Kashdan, T. B., biswas-Diener, R., & King, L.A. (2008). Reconsidering happiness: The cost of distinguish between hedonics and eudaimonia. *The Journal of Positive Psychology, 3*(4), 219-233.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73*(3), 539-548.

Keyes, C. L. M. (1998). Social well-being. *Social Psychology Quarterly*, 121-140.

Keyes, C.L.M. (2013). *Mental Well-Being: International Contributions to the Study of Positive Mental Health*. 7th ed. Springer, Dordrecht, The Netherlands.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*(2), 212-228.

Kim, J. E., & Moen, P. (2002). Retirement Transitions, Gender, and Psychological Well-Being A Life-Course, Ecological Model. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *57*(3), P212-P222.

King, L. A. (2001). The health benefits of writing about life goals. *Personality and Social Psychology Bulletin*, *27*(7), 798-807.

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford press.

Lalive, R., & Stutzer, A. (2010). Approval of equal rights and gender differences in well-being. *Journal of Population Economics*, *23*(3), 933-962.

Langeland E. *(*2007*) Sense of Coherence and Life Satisfaction in People Suffering from Mental Health Problems. An Intervention Study in Talk-Therapy Groups with Focus on Salutogenesis. (PhD thesis).* University of Bergen, Bergen.

Langeland, E., Riise, T., Hanestad, B. R., Nortvedt, M. W., Kristoffersen, K., & Wahl, A. K. (2006). The effect of salutogenic treatment principles on coping with mental health problems: A randomised controlled trial. *Patient Education and Counseling*, *62*(2), 212-219.

Layard, R., Mayraz, G., & Nickell, S. (2010). Does relative income matter? Are the critics right. *International differences in well-being*, *28*, 139-66.

Lewandowski Jr, G. W. (2009). Promoting positive emotions following relationship dissolution through writing. *The Journal of Positive Psychology*, *4*(1), 21-31.

Linley, P. A., & Joseph, S. (Eds.). (2004). *Positive psychology in practice*. 10th ed. Wiley, Hoboken, NJ.

Lloyd, K., & Devine, P. (2012). Psychometric properties of the Warwick-Edinburgh mental well-being scale (WEMWBS) in Northern Ireland. *Journal of Mental Health*, *21*(3), 257-263.

López, M. A., Gabilondo, A., Codony, M., García-Forero, C., Vilagut, G., Castellví, P., ... & Alonso, J. (2012). Adaptation into Spanish of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) and preliminary validation in a student sample. *Quality of Life Research*, 1-6.

Lopez, S. J., & Snyder, C. R. (Eds) (2011). *The Oxford handbook of positive psychology.* (2nd ed,) Oxford University Press, Oxford

Lyubomirsky, S. (2008). *The how of happiness: A scientific approach to getting the life you want.* Penguin Press, New York.

Lyubomirsky, S., King, L., & Diener, E. (2005). The benefits of frequent positive affect: Does happiness lead to success?. *Psychological Bullettin*, *131*(6), 803-855.

Maheswaran, H., Weich, S., Powell, J., & Stewart-Brown, S. (2012). Evaluating the responsiveness of the Warwiwck Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health and Quality of Life Outcomes, 10*(1), 156-164.

Manzi, C., Vignoles, V. L., Regalia, C., & Scabini, E. (2006). Cohesion and Enmeshment Revisited: Differentiation, Identity, and Well-Being in Two European Cultures. *Journal of Marriage and Family, 68*(3), 673-689.

Marchante, A. J., Ortega, B., & Sánchez, J. (2006). The evolution of Well-Being in Spain (1980-2001): a regional analysis. *Social Indicators Research, 76*(2), 283-316.

McBride, M. (2001). Relative-income effects on subjective well-being in the cross-section. *Journal of Economic Behavior & Organization, 45*(3), 251-278.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568-592.

Menzies, V. (2000). Depression in schizophrenia: Nursing care as a generalized resistance resource. *Issues in Mental Health Nursing, 21*(6), 605-617.

Milberg, A., & Strang, P. (2007). What to do when 'there is nothing more to do'? A study within a salutogenic framework of family members' experience of palliative home care staff. *Psycho-Oncology, 16*(8), 741-751.

Mindell, J., Biddulph, J. P., Hirani, V., Stamatakis, E., Craig, R., Nunn, S., & Shelton, N. (2012). Cohort profile: the health survey for England. *International Journal of Epidemiology, 41*(6), 1585-1593.

Mitchell, J., Stanimirovic, R., Klein, B., & Vella-Brodrick, D. (2009). A randomised controlled trial of a self-guided internet intervention promoting well-being. *Computers in Human Behavior, 25*(3), 749-760.

Morgan, A., & Ziglio, E. (2007). Revitalising the evidence base for public health: an assets model. *Promotion & Education, 14*(2 suppl), 17-22.

Muthén, L.K. and Muthén, B.O. (1998-2012). Mplus User's Guide. Seventh Edition. Muthén & Muthén, Los Angeles, CA.

NatCen Social Research (2011). *Health Survey for England 2011. User Guide*. Department of Epidemiology and Public Health, University College, London.

Newbigging, k., Bola, M., & Shah, A. (2008). *Scoping exercise with Black and minority ethnic groups on perceptions of mental wellbeing in Scotland.* Retrieved October 2, 2013, from http://www.healthscotland.com/uploads/documents/7977-RE026FinalReport0708.pdf

NHS Health Scotland (2007). Health Education Population Survey: Update from 2006 survey. *Glasgow: NHS Health Scotland.* UK Data Archive, Colchester.

Nie, N. H., Bent, D. H., & Hull, C. H. (1975). *SPSS: Statistical package for the social sciences* (Vol. 227). McGraw-Hill,New York.

Nolen-Hoeksema, S., & Rusting, C. L. (2003). Gender Differences in Well-Being. *Well-being: The foundations of Hedonic Psychology*, 330-352.

Nunnally, J. C., & Bernstein, I. H. (1978). *Psychometric testing*. McGraw-Hill, San Francisco, CA.

Odou, N., & Vella-Brodrick, D.A. (2013). The efficacy of positive psychology interventions to increase well-being and the role of mental imagery ability. *Social Indicators Research, 110*(1), 111-129.

Oostendorp, R. H. (2009). Globalization and the gender wage gap. *The World Bank Economic Review*, *23*(1), 141-161.

Organisation for Economic Co-operation and Development. (2013). *OECD Economic Surveys*. OECD Pub.

Otake, K., Shimai, S., Tanaka-Matsumi, J., Otsui, K., & Frederickson, B. L. (2006). Happy people become happier through kindness: A counting kindness intervention. *Journal of Happiness Studies*, *7*(3), 361-375.

Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, *46*(1), 1-18.

Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, *50*(302), 157-175.

Peters, M. L., Flink, I. K., & Linton, S. J. (2010). Manipulating optimism: Can imagining a best possible self be used to increase positive future expectancies?. *The Journal of Positive Psychology*, *5*(3), 204-211.

Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health?. *Psychological Bulletin*, *131*(6), 925-971.

Ragonesi, M. (2012). *Mental Well-Being and the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): An analysis of its psychometric performance in screening postnatal depression*. Thesis for the Master in Clinical Psychology. University of Bologna.

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.

Reed, G. L., & Enright, R. D. (2006). The effects of forgiveness therapy on depression, anxiety and posttraumatic stress for women after spousal emotional abuse. *Journal of Consulting and Clinical Psychology, 74*(5), 920-929.

Ruini, C., Ottolini, F., Rafanelli, C., Ryff, C. D., & Fava, G. A. (2003). La validazione italiana delle Psychological Well-being Scales (PWB). *Rivista di psichiatria*, *38*(3), 117-130.

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

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, *69*(4), 719-729.

Sacks, D. W., Stevenson, B., & Wolfers, J. (2010). *Subjective well-being, income, economic development and growth* (No. w16441). National Bureau of Economic Research.

Savalei, V., & Bentler, P. M. (2006). Structural equation modeling. In R. Grover & M. Vriens (Eds.), The handbook of marketing research: Uses, misuses, and future advances (pp. 330-364). Thousand Oaks, CA: Sage.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Schmitt, M. T., Branscombe, N. R., Kobrynowicz, D., & Owen, S. (2002). Perceiving discrimination against one's gender group has different implications for well-being in women and men. *Personality and Social Psychology Bulletin*,*28*(2), 197-210.

Schmolke, M. (2003). Discovering positive health in the schizophrenic patient: A salutogenetic-oriented study. *Verhaltenstherapie, 13*(2), 102-109.

Schueller, S. M. (2010). Preferences for positive psychology exercises. *The Journal of Positive Psychology, 5*(3), 192-203.

Schutte, N. S., Malouff, J. M., Hall, L. E., Haggerty, D. J., Cooper, J. T., Golden, C. J., & Dornheim, L. (1998). Develpment and validation of a measure of emotional intelligence. *Personality and Individual Differences, 25*(2), 167-177.

Scottish Governement. (2009). *Towards a mentally flourishing Scotland: 2009-2011 Policy and Action Plan*. Retrieved September 30, 2013, from http://www.scotland.gov.uk/Publications/2009/05/06154655/0

Seligman, M. E. (2002). *Authentic happiness: Using the new positive psychology to realize your potential for lasting fulfillment*. New York: Free Press/Simon and Schuster.

Seligman, M. E. (2012). *Flourish: A visionary new understanding of happiness and well-being.* New York:Simon and Schuster.

Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: an introduction. *American Psychologist, 55*(1), 5-14.

Seligman, M. E., Steen, T. A., Park, N., & Peterson, C. (2005). Positive psychology progress: empirical validation of interventions. *American Psychologist, 60(5), 410-421*.

Sergeant, S., & Mongrain, M. (2011). Are positive psychology exercises helpful for people with depressive personality styles?. *The Journal of Positive Psychology, 6*(4), 260-272.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591-611.

Sheldon, K. M., & Lyubomirsky, S. (2006). How to increase and sustain positive emotion: The effects of expressing gratitude and visualizing best possible selves. *The Journal of Positive Psychology, 1*(2), 73-82.

Silberman, J. (2007). Positive intervention self-selection: Developing models of what works for whom. *International Coaching Psychology Review, 2*(1), 70-77.

Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: A practice-friendly meta-analysis. *Journal of Clinical Psychology, 65*(5), 467-487.

Sixsmith, A., & Sixsmith, J. (2008). Ageing in place in the United Kingdom. *Ageing International*, *32*(3), 219-235.

Skärsäter, I., Langius, A., Ågren, H., Häggström, L., & Dencker, K. (2005). Sense of coherence and social support in relation to recovery in first-episode patients with major depression: A one-year prospective study. *International Journal of Mental Health Nursing, 14*(4), 258-264.

Snyder, C. R., Lopez, S. J., & Pedrotti, J. T. (Eds.). (2010). *Positive psychology: The scientific and practical explorations of human strengths*. Newbury Park, CA: Sage Publications.

Sörbom, D. (1989). Model modification. *Psychometrika, 54*(3), 371-384.

Steiger, J. H., & Lind, J. C. (1980, May). Statistically based tests for the number of common factors. In *Annual meeting of the Psychometric Society, Iowa City, IA* (Vol. 758).

Stewart-Brown, S. (2002). Measuring the parts most measures do not reach. *Journal of Mental Health Promotion, 1*, 4-9.

Stewart-Brown, S. (2013). The Warwick-Edimburgh Mental Well-Being Scale (WEMWBS): Performance in Different Cultural and Geographical Groups. In C.L.M. Keyes (2013), *Mental Well-Being: International Contributions to the Study of Positive Mental Health.* (pp. 133-150). Dordrecht, Netherlands: Springer.

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkison, J. & Weich, S. (2009) Internal Construct Validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes*, *7*(1), 15-22.

Streiner, D. L., & Norman, G. R. (2008). *Health measurement scales: a practical guide to their development and use*. Oxford, Oxford University Press.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using Multivariate Statistics: International Edition*. Boston: Allyn & Bacon.

Taggart, F., Friede, T., Weich, S., Clarke, A., Johnson, M., & Stewart-Brown, S. (2013). Cross cultural evaluation of the Warwick-Edinburgh mental well-being scale (WEMWBS)- a mixed methods study. *Health and Quality of Life Outcomes*, *11*(27), (1474-7525).

Tennant, R ., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., ... & Stewart-Brown, S. (2007a). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, *5*(1), 63-76.

Tennant, R., Fishwick, R., Platt, S., Joseph, S., & Stewart-Brown, S. (2006). *Monitoring positive mental health in Scotland: validating the Affectometer 2 scale and developing the Warwick-Edinburgh Mental Well-being Scale for the UK*. Edinburgh, NHS Health Scotland.

Tennant, R., Joseph, S., & Stewart-Brown, S. (2007b). The Affectometer 2: a measure of positive mental health in UK population. *Quality of Life Research, 16*(4), 687-695.

Ullman, J. B., & Bentler, P. M. (2001). Structural equation modeling. In B.G. Tabachnick & L.S. Fidell (2012).*Using multivariate statistics*. Boston: Allyn & Bacon

Van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, *9*(4), 486-492.

Waiyavutti, C., Johnson, W., & Deary, I. J. (2012). Do personality scale items function differently in people with high and low IQ?. *Psychological Assessment, 24*(3), 545-555.

Wallerstein, N. (2006). *What is the evidence on effectiveness of empowerment to improve health*. Copenhagen: WHO Regional Office for Europe, n. 37.

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), 1063-1070.

Waugh, C. E., & Fredrickson, B. L. (2006). Nice to know you: Positive emotions, self-other overlap, and complex understanding in the formation of a new relationship. *Journal of Positive Psychology*, *1*(2), 93-106.

Weichselbaumer, D., & Winter-Ebmer, R. (2005). A Meta-Analysis of the International Gender Wage Gap. *Journal of Economic Surveys*, *19*(3), 479-511.

Weissbecker, I., Salmon, P., Studts, J. L., Floyd, A. R., Dedert, E. A., & Sephton, S. E. (2002). Mindfulness-based stress reduction and sense of coherence among women with fibromyalgia. *Journal of Clinical Psychology in Medical Settings*, *9*(4), 297-307.

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*(5), 719-751.

WHO (2001). *Strengthening mental health promotion*. Geneva, World Health Organization (Fact sheet, No. 220).

Williams, G. C., Lynch, M. F., McGregor, H. A., Ryan, R. M., Sharp, D., & Deci, E. L. (2006). Validation of the" Important Other" Climate Questionnaire: Assessing Autonomy Support for Health-Related Change. *Families, Systems, & Health*, *24*(2), 179-194.

Williams, S., & Shiaw, W. T. (1999). Mood and organizational citizenship behavior: The effects of positive affect on employee organizational citizenship behavior intentions. *The Journal of Psychology*, *133*(6), 656-668.

Wissing, M. P., & Temane, Q. M. (2013). The prevalence of levels of well-being revisited in an African context. In C.L.M. Keyes (2013), *Mental Well-Being: International Contributions to the Study of Positive Mental Health.* (pp. 71-90). Dordrecht, Netherlands, Springer.

World Health Organization (2001). *Mental Health: New Understanding, New Hope.* World Health Report, Geneva, Switzerland.

World Health Organization. (1947). *Constitution of the World Health Organization: Signed at the International Health Conference, New York, 22 July 1946.* World Health Organization, Interim Commission.