

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA

SEDE DI CESENA

FACOLTÀ DI PSICOLOGIA

CORSO DI LAUREA MAGISTRALE IN

PSICOLOGIA CLINICA

Performance of the Warwick-Edinburgh Mental Well-Being Scale
(WEMWBS) as a screening tool for depression in UK and Italy

Tesi di laurea in

Tecniche di Valutazione Testistica in Psicologia Clinica

Relatore

Presentata da

Prof. Paola Gremigni

Donatella Bianco

Sessione II

Anno Accademico 2011/2012

INDEX

INTRODUCTION.....	1
1. THE CONSTRUCTS OF MENTAL WELL-BEING, PSYCHOLOGICAL DISTRESS AND DEPRESSION.....	5
I. Mental well-being.....	5
i. The Hedonic approach.....	6
ii. The Eudaimonic approach.....	8
iii. Toward integration between Hedonic and Eudaimonic Approaches.....	11
II. Depression and Psychological Distress.....	12
III. Positive <i>versus</i> negative wording of items.....	17
2. LITERATURE REVIEW.....	21
I. Validation studies of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS).....	21
II. Validation studies of the Center for Epidemiologic Studies Depression scale (CES-D).....	24
i. Cross-cultural validation studies.....	25
ii. Specificity and sensitivity of the CES-D.....	27
a. Non-clinical population.....	27
b. Comparison of Clinical and Non-clinical Populations.....	29
c. Clinical population.....	29
III. Conclusion.....	31
3. THIS STUDY.....	33
I. Introduction and objectives.....	33
II. Procedures and participants.....	34

III.	Measures.....	36
IV.	Statistical analysis.....	38
	i. Statistical analyses on the UK sample.....	38
	ii. Statistical analyses on the Italian sample.....	39
4.	RESULTS.....	43
	I. Results on the UK sample.....	43
	II. Results on the Italian sample.....	50
	i. ANOVA analyses' results for 44.5 cut-off point.....	50
	ii. ANOVA analyses' results for 40.5 cut-off point.....	53
5.	DISCUSSION.....	57
6.	CONCLUSIONS.....	63

INTRODUCTION

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) is a self-report questionnaire developed for measuring psychological well-being at a community level (Tennant, Hiller, Fishwick, Platt, Joseph, Weich, Parkinson, Secker, & Stewart-Brown, 2007). In particular, it refers to a double concept of mental well-being as a hedonic and eudemonic construct, providing information on positive affect, satisfying interpersonal relationships and positive personal functioning. Since its first validation, several studies have been conducted to investigate the psychometric properties of this questionnaire and, on the whole, good reliability and validity were confirmed.

In the last years, a new demand has arisen from primary care and social field to have questionnaires that are highly acceptable by the users. The WEMWBS is considered a useful tool for the screening and monitoring of mental well-being in particular because of its positively worded items, which make the questionnaire acceptable by users. For this reason, a cut-off score of this scale has been required by social workers and helplines in UK to be used for the screening of depression.

According to the Mental Health Foundation's statistics, in UK 1 in 4 people will experience some kind of mental health problem in one year and depression, mixed with anxiety, is the most common mental disease, with almost 9% of people meeting criteria for diagnosis. In particular, between 8-12% of the population experience depression in any year and 1 in 5 older individuals is affected by this disorder. In compliance with the statistics in Italy (Associazione per la Ricerca sulla Depressione), almost 1.5 million people were affected by depression in 2008 while 10% of the Italian population experienced almost one episode of depression. The World Health Organization identified the unipolar depressive disorders as the third leading cause of the global burden of disease in 2004 and anticipated it will be the first by 2030: at present, they estimated that 350 million people globally suffer from depression.

In view of the picture described above, methods for the screening of mental illness are increasingly required: several studies investigate the utility of screening programs (e.g., Sharp & Lipsky, 2002; Kerr, 2001; Valenstein, Vijan, Zaber, Bohem, Butter,

2001) and found that brief self-report tools are actually useful for the screening of people at risk of depression in primary care center; furthermore, it was found that the use of these instruments could maximize the cost-effectiveness of the screening programs in presence of effective treatment.

Screening the general population may allow identify those people who are at risk of developing mental diseases. Such an action may lead to more than one benefit: it can help improving mental health through preventive programs on the community, reducing the risk of pain and psychological illness, and restraining social costs due to the management of full-blown diseases.

This study arose from the will of the Department of Mental Health of Warwick University to face this challenge and add a little contribution to the research of new instruments for the screening of depression.

Our main aim was the investigation of possible WEMWBS cut-off points for the screening of this mental disorder, since we believed that this scale has the potential characteristics required of a screening tool. In particular, we expected to find one cut-off point for the screening of psychological distress indicating the need for further assessment and one more restrictive cut-off for the detection of Major Depression.

In order to achieve this goal, we decided to compare scores on the WEMWBS with scores on the Center for Epidemiologic Studies Depression Scale (CES-D), because of the common utilization of the latter for the screening of depression in the epidemiologic studies. We planned to obtain, in this way, scores on the WEMWBS corresponding to the CES-D cut-off points used to screen probable and possible depression cases, respectively.

Finally, in collaboration with the University of Bologna, we wanted to test if the cut-off points identified on the UK sample well screened between depressed and non-depressed individuals also in an Italian sample. To achieve this goal, the Italian sample was dichotomized in depressed and non-depressed subjects on the basis of the two cut-off points identified in the first part of the research; subsequently, the groups obtained were compared in the level of depression and psychological distress as measured by two other instruments, the Psychological General Well-Being Index (PGWBI) and the Well-Being Index (WHO-5). In this way we wanted to verify if the differences between potentially depressed and non-depressed subjects in the scores

on these two questionnaires were statistically significant or not and, accordingly, if the cut-offs detected on the UK sample were reliable and appropriate also in this case.

In order to achieve its goals, this study was structured as follows.

The *first chapter* introduces to the concept of mental well-being and its potential link with opposite constructs such as psychological distress and depression. Follow a brief literature review on the WEMWBS and the CES-D where we analyze the main studies that validated the WEMWBS and the CES-D and report a brief history about the development of these tools, their psychometric properties and summary of the main cross-cultural studies. As regards the CES-D, the issue about the optimal cut-off point for the screening of depressed people is reported and studies which investigated this matter are analyzed in the subsequent sub-paragraphs: the first of these considers researches conducted on non-clinical population (community samples); the second one regards studies which compare clinical and non-clinical populations; the last sub-paragraph contains the studies about the research of cut-off point on the clinical populations. The third paragraph of this chapter reports the conclusions drawn from the previous two sections and the choice of the CES-D cut-off points more appropriate for their use as gold-standard.

The *second chapter* summarizes the methods followed to investigate the hypotheses of this study. In the first paragraph, a description of the procedures used to realize the study and the characteristics of the participants is reported; in particular, in the first sub-paragraph, a description of the study whence the UK sample was drawn is done and baseline characteristics are described. In the second one, method of recruitment and characteristics of the Italian sample are depicted.

The second paragraph of the chapter describes the tools utilized.

The third paragraph shows the statistical analyses used in the study. A first sub-paragraph summarizes the analyses conducted on the UK sample in order to obtain two reliable cut-off points for the screening of probable and possible cases of depression respectively. The method used to investigate the correlation between the WEMWBS and the CES-D scores is reported as well. The last sub-paragraph states the same contents about the second part of the study conducted on the Italian sample: the dichotomization of the sample and the comparison between the groups obtained

in the PGWBI and the WHO-5 scores are described; this procedure allows to test if the cut-off points previously obtained are appropriate for screening purposes also in this sample.

The *third chapter* describes the results obtained in the study on the UK sample (first paragraph) and on the Italian one (second sub-paragraph), using summary tables and graphs for a better and easier comprehension.

The *fourth chapter* is reserved to an interpretation of the results and to a comparison of them with knowledge drawn from the literature. A summary of the hypotheses and the methods used to investigate them is done. After summarizing the results obtained and their significance, a comparison across Italian and UK samples is shown to verify the initial hypotheses. Limitations of the study are then discussed; several remarks on the tools utilized and the results obtained are included in this section as well.

The *fifth chapter* reports the conclusions of our study. An overall review of the initial aims and the reasons that urged to realize the research is provided. The usefulness of findings is discussed in terms of both knowledge added and clinical applicability. The chapter is concluded with some considerations and ideas for possible future researches.

1. THE CONSTRUCTS OF MENTAL WELL-BEING, PSYCHOLOGICAL DISTRESS AND DEPRESSION

I. Mental well-being

The scientific literature neglected for long time the construct of mental well-being. Traditionally mental health research was focused on psychological impairment and illness, traducing mental well-being as lack of illness rather than presence of wellness (Ryff & Singer, 1996).

Around 1960, the world of psychiatry and mental health attended a shift in interest toward new topics such as promotion of mental well-being (Diener, 1984), growth (Deci & Ryan, 1975) and wellness (Cowen, 1991). This drift was anticipated by the World Health Organization that in 1948 described health as “a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity” (WHO, 1948). Therefore, a more specific definition of positive mental health was provided and described as a state “which allows individuals to realize their abilities, cope with the normal stresses of life, work productively and fruitfully, and make a contribution to their community” (WHO, 2004) and as “the capacity for mutually satisfying and enduring relationships” (WHO, 2001). As cited by Clarke (Clarke, Friede, Putz, Ashdown, Martin, Blake, Adi, Parkinson, Flynn, Platt, Stewart-Brown, 2011), the Royal Society in UK also defined well-being as a “positive and sustainable mental state that allows individuals, groups and nations to thrive and flourish”.

This change was fostered by interesting discoveries that raised new clinical and methodological questions; for example, Bradburn (Bradburn, 1969) found that positive and negative affect were not just opposites of one another; the author thought that these two affects must be studied separately since they are independent and have different correlates. Accordingly, clinical efforts to eliminate or reduce metal impairment and distress could not correspond to positive states: reducing depressive or anxiety symptoms could not lead to happiness or wellness.

Otherwise, the tendency to “thriving” and “flourishing” never had a one-dimensional definition: different concepts were taken into consideration to explain it. Otherwise,

the great amount of theoretical contributions can be summarized in two main philosophies: the hedonic and the eudaimonic approaches. These are clearly explained by Ryan and Deci (Ryan and Deci, 2001), thus, we start from their review to investigate the construct of mental well-being.

i. The Hedonic approach

Hedonism is a school of thought with ancient origins. It is supposed to be born with Aristippus, student of Socrates and founder of the Cyrenaic school. He taught that the only good is pleasure intended as positive gratifying sensations, in particular the momentary ones.

Other philosophers subsequently adhered to this thought; for example, De Sade and Hobbes taught that the ultimate goal of life is the pursuit of happiness and pleasure. Similarly, in 18th-19th century, Jeremy Bentham fed the Utilitarian theory; on his opinion, every action should be aimed to the realization of the “Hedonic Calculus”, the greatest possible amount of happiness and pleasure for the individual and for the community.

Subsequently, several psychologists took cue from this original concept of Hedonism and developed theories on human well-being gradually shifting from bodily pleasures to a broader concept of happiness (e.g., Diener, 1984; Kahneman, 1999).

The most notable theorists in this field described well-being as consisting of three main components: life satisfaction, positive mood, and absence of negative mood or happiness (Diener, 1984; Bradburn, 1969). In particular, the focus was on cognitive and affective side of well-being: on their opinion, it is possible to investigate an individual’s mental well-being status asking for *his* personal perception of satisfaction with life: as Publius Syrius said, “The happy man is not he who seems thus to others, but who seems thus to himself”. This view emphasizes people’s opinion about their life over an expert evaluation; for this reason well-being based on a hedonic theoretical substrate has been defined subjective well-being (SWB).

The affective side of this construct was theorized by Bradburn who described positive mental well-being as the preponderance of pleasant emotional experiences on negative affect (Bradburn, 1969); in order to measure well-being in this meaning, Bradburn developed the Affect Balance Scale (ABS), a questionnaire that provides

the individual's affect equilibrium measuring both positive affect and negative affect as independent variables. An example of ABS item is "I felt that things were going my way". Other scales were then developed to measure the affective well-being, such as the Positive and Negative Affect Scale (PANAS) (Watson, Clark, Tellegen, 1988); this questionnaire confirmed the independence of positive and negative affects giving two different scores of them and showed good psychometric properties (better than the ABS ones).

On the other hand, Diener was one of the main authors that investigated the SWB particularly as a cognitive evaluation of overall life satisfaction (Diener, 1984). He developed the Satisfaction with Life Scale whose total score depicted the individual's level of SWB. Items were such as "In most ways my life is close to my ideal", "I'm satisfied with my life", "The conditions of my life are excellent", and it required an evaluation of agree/disagree. The questionnaire allowed both a general score of SWB and an investigation of the different domains analyzed (e.g. relationships, family).

According with SWB concept, several theorists tried to depict the most positive domains in one's life and subsequently developed psychological instruments for measuring mental well-being. For example, Rosemberg focused on self-esteem as a construct strongly relevant to SWB (Rosemberg, Schooler, Schoenbach, Rosemberg, 1995); Scheier and Carver considered the benefits of positive thinking and optimism (Scheier & Carver, 1993) while Antonovsky, with his "salutogenic model", analyzed the relationship between health, stress, and coping and stated that well-being depends on the experience of a strong "sense of coherence" (Antonovsky, 1993).

In a recent review of Diener and colleagues (Diener, Lucas, Oishi, 2009), three theoretical approaches of happiness at the bottom of SWB were distinguished. The first of them focused on the hypothesis that reduction of tensions leads to happiness: from this point of view, an individual achieve high level of SWB when move toward a state that is significant for himself. The second approach is centered on process and activity and is directly related to the first one: if a person has a goal, he needs to activate himself and be energetic in order to fulfill his aim and, consequently, achieve a good level of SWB. The third theory on SWB focuses on genetic and personality predisposition. Diener found that, although there are differences in the degree of SWB when measured in few occasions, this pattern of change tends to

become more stable if the measurement is conducted in many situations (Diener & Larsen, 1984).

This emphasis on the genetic set-point was also confirmed by Lyubomirsky in her work “Pursuing happiness: The architecture of sustainable change” (Lyubomirsky, Sheldon, Schkade, 2005). The author supported the idea that personal well-being is the outcome of three main ingredients, the main of which is the genetic one: 50% of the contribution for the development of happiness is done by genetic set-point, 40% by happiness-relevant activities and practices while 10% is adduced by happiness-relevant circumstantial factors. According to this view, Diener stated that two main personality traits are distinguishable as strictly related with SWB (Diener & Lucas, 1999), the extraversion and the neuroticism; in particular, the first one showed to be correlated moderately to strongly with pleasant affect while the second one resulted strongly correlated with negative affect.

Otherwise, the strong genetic contribution doesn't mean that environment and circumstances of life have not influence on individuals' happiness and, to this end, several researchers have recently investigated life events significant for SWB. Diener found that demographic factors such as age, sex, income, marital status and religiosity are relevant for SWB, (although their effects are not too large) but their relevance depends on people's aims, personality and, most of all, culture (Diener et al., 2009). About this last factor, it has been confirmed, for example, that unmarried individuals who live together are happier in individualistic culture and happiest in collectivistic one (Diener, Gohm, Suh, Oishi, 2000) or that having a congruence personality leads to SWB in the individualistic culture and less in the collectivistic one (Suh, 1999). As stated by Diener, “cultural norms can change the correlates of subjective well-being” (Diener et al., 2009).

ii. The Eudaimonic approach

A careful review of Aristotle's “*Nicomachean Ethics*” conducted by Carol Ryff (Ryff & Singer, 2008), clearly explains the basic principles of the development of the concept of Psychological Well-Being (PWB). Whereas the Hedonic approach emphasizes pleasure and happiness as necessary conditions for developing well-being, the Eudaimonic view focuses on the development of one's potential and on

self-realization as path to mental well-being (Ryff, 1989). Aristotle, in his attempt to answer the universal question on how an individual should live, warns people against the risk of conducting “life suitable to beasts” (consisting of pleasures and appetites as the only goals to achieve) and teaches how to get the true self-realization: in this philosophic framework, the highest human good becomes the “activity of the soul in accordance with virtue, and if there be more than one virtue, in accordance with the best and most complete”. As well recently interpreted and explained by the heir to the Hellenic school David Norton (1976), each individual has to live in order to realize his *daimon*, depiction of innate potentialities of each one.

Several authors, including Mill (1843), Russell (1930) and Fromm (1981), drew their inspiration from Aristotle’s principles to elaborate new theories on the meaning of human life, providing a broad substrate for the subsequent psychological theories on PWB. In particular, Ryff confirmed she was guided in her training by three main schools of thought: the developmental psychology, the existential and humanistic psychology and the clinical psychology. About the contribution of each of these view to the development of the concept of PWB, the first one (including researchers such as Buhler and Erikson) emphasizes the developmental challenges that each individual has to deal in his path to human growth; existential psychologists (Allport, Rogers, Maslow) teach that courage and hardiness to persist in front of life difficulties are the necessary tools to achieve authentic being while clinical psychologists such as Jung and Jahoda gave a new definition of mental health not only focused on dysfunction but described in positive terms (Ryff, 2008).

An integration of these approaches and theories resulted in Carol Ryff’s six-factor model of psychological well-being (Ryff, 1982, 1985, 1989). The factors considered depicted the core dimensions of well-being; a brief description of them follows:

- *Autonomy*: it involves one’s independence and strong self-confidence. People with high level of autonomy are inclined to regulate their behavior and acts on the basis of internal and personal standards and are not prone to think or act as social pressure suggests. As highlighted by the author, autonomy is the most “western” of all of the following dimensions, in keeping with an individualistic life philosophy.

- *Self-acceptance*: it is described as a central feature of mental health as well as sign of maturity, good functioning and self-actualization. An individual with significant level of self-acceptance knows and accepts himself, his strengths and weaknesses with mature awareness. Generally, he looks at past with positive mood, considering each life events as important or functional for his growth and development.
- *Environmental mastery*: this dimension is strictly related to Allport's "extend the self". It implies the capacity to manage and control surrounding environment and to change it with physical or mental activities in order to achieve an environment more suitable to their needs and being. These persons are able to catch each useful occasion that the surrounding context offers.
- *Purpose in life*: people with strong purpose in life know what they want and activate themselves to achieve their aims. Creating meaning and direction is their main challenge and a positive evaluation of past and current life is a distinctive characteristic of their thinking. Sense of directedness and intentionality are key-words to well comprehend the meaning of this dimension.
- *Personal growth*: it is the depiction of the eudaimonic concept of self-realization. People perceive their life as ever-increasing and ever-growing and are open to new experiences that could be useful to the realization of their own potential. They look at the future with positive mood and dynamic attitude.
- *Positive relations with others*: this dimension focuses on the primacy of love, affection and empathy. These people are able to share emotions and affection and to create intimate relationships. This attitude seems to be universal and constitutes a key-feature of well-being and satisfactory life (Ryff & Singer, 1998).

A self-report questionnaire was developed by Carol Ryff for measuring PWB through the investigation of these six dimensions: the PWB Scale consists of 84 items (but other versions by 42 and 18 items exist) and 6 subscales, one for each dimension. Confirmatory factorial analyses conducted by the authors and other

researchers confirmed the stability of the six-factor model (Cheung & Chang, 2005; Clarke, Marshall, Ryff, Wheaton, 2001; Ryff and Keyes, 1995; Springer and Hauser, 2006). An Italian version of the scale was also developed (Ruini, Ottolini, Rafanelli, Ryff, Fava, 2003), confirming good psychometric properties of the scale that showed to be able to provide a complete evaluation of PWB also in Italy.

Self-Determination Theory (SDT) is another theoretical approach whose core concept is the eudaimonic well-being (Ryan & Deci, 2001). The authors state that the only way to achieve mental health and psychological growth is fulfilling three main human needs: Autonomy, Competence and Relatedness. This theory is similar to Ryff's one: they agree on the content of being eudaimonic and on the full realization of one's potential as condition for the development of psychological well-being. Otherwise, if Ryff's model forecasts that the six dimensions define well-being, Ryan and Deci state that their factors only foster it. Furthermore, the SDT model assume that satisfaction of the basic psychological needs lead to the achievement of SWB as well as PWB; so SWB is considered one of several indicators of well-being, although the authors clearly distinguish the two constructs, their causes and characteristics.

iii. Toward integration between Hedonic and Eudaimonic approaches

A study conducted in 2003 in the United States analyzed the encounter of the Hedonic and the Eudaimonic traditions in order to achieve integration between these two approaches (Keyes, Shmotkin, Ryff, 2003). In particular, it was hypothesized that hedonic and eudaimonic well-beings were "conceptually related but empirically distinct"; therefore it was supposed that different combinations of these two perspectives relate differentially to socio-demographic factors and personality.

In order to investigate these aims, a sample composed of 3.032 Americans aged 25-74 was recruited, SWB was measured by a single-item measure of global life satisfaction and scales of positive and negative affect while level of PWB was investigated using the Psychological Well-Being Scale.

Results of the study can be summarized as follows.

The main hypothesis was confirmed: results suggested that the best fitting model is composed of two main correlated latent dimensions, the SWB and the PWB, and

that, although they resulted to be correlated, each one depicted a different facet of the same construct.

About the relationships between well-being and socio-demographic factors and personality traits, interesting results were achieved. Briefly, individuals with high level of both SWB and PWB seemed to have higher level of education and an age range comprised between midlife and older age; these people were characterized by high extraversion and high conscientiousness. On the other hand, low level of both SWB and PWB were characteristics of younger adults with low level of education and low extraversion and conscientiousness. When analyzing off-diagonal type of well-being, authors found that people reporting high SWB and low PWB were mostly midlife or older adults with less education and low level of openness to new experiences; conversely, individuals with high PWB and low SWB were younger and characterized by high level of education and high openness to experience, neuroticism and conscientiousness.

We could say that a good equilibrium between SWB and PWB requires high level of education, extraversion and conscientiousness, older age and less neuroticism. An equivalent level of the two types of well-being is extremely positive particularly because of their complementary relationship: high level of both SWB and PWB provides a “sense of self-congruency”.

On the other hand, when SWB and PWB exist at different levels, they may compensate each other and try to maximize the well-being status achievable with the available resources.

II. Depression and Psychological Distress

Consistent with the DSM-IV-TR definition (American Psychiatric Association, 2000), Major Depressive Disorder is characterized by the presence of one or more Major Depressive Episode. This latter is described by DSM criteria as follows:

A) Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

- 1) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). Note: In children and adolescents, can be irritable mood.
 - 2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation made by others).
 - 3) Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. Note: In children, consider failure to make expected weight gains.
 - 4) Insomnia or hypersomnia nearly every day.
 - 5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
 - 6) Fatigue or loss of energy nearly every day.
 - 7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
 - 8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
 - 9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.
- B) The symptoms do not meet criteria for a Mixed Episode.
- C) The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
- D) The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).
- E) The symptoms are not better accounted for by Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation.

On the basis of number of episodes, it is possible to diagnose a single-episode Major Depressive Disorder or a recurrent disorder (2 or more episodes).

The concept of psychological distress is frequently used in clinical practice but there are not specific criteria for its definition.

Massé, in his attempt to develop a scale for the measurement of this construct, conducted a qualitative and a quantitative study on a French sample, the first of which led the author to the detection of six specific features related to the psychological distress: demoralization and pessimism, anguish and stress, self-depreciation, social withdrawal and isolation, somatization, withdrawal into oneself (Massé, 2000).

Afterwards, Ridner, in her work “Psychological distress: concept analysis”, identified 5 main dimensions characterizing this mental status (Ridner, 2004):

1. Perceived inability to cope, involving hopelessness, avoidance of issues, inability to make decision without others’ support.
2. Change in emotional status, namely depression, anxiety, irritableness, withdrawal from others.
3. Discomfort, such as sadness, pain, anger, hostility.
4. Verbal and physical communication of discomfort, such as expressing lack of hope for future or restlessness and avoidance of eye contact.
5. Harm, involving features such as pain and change in vital signs.

As we can see, a common feature that recurs in both the studies is depression, described in different meanings and retrievable in the DSM criteria previously listed. It has to do with depressive symptoms that although they don’t meet the criteria for a full-blown disorder, still compromise individual’s daily life and his mental well-being.

Several studies attempted to investigate the relationship between psychological distress/depressive symptoms and depressive disorders, with preference for Major Depressive Disorder. Among them, the studies that analyzed the prodromal symptoms of depression resulted to be particularly useful for investigating this relationship (e.g., Fava, Grandi, Canestrari, Molnar, 1990; Fava, Grandi, Zilezny, Canestrari, Morphy, 1994; Mahnert, Reicher, Zalaudek, Zapotoczky, 1997).

A prodromal symptom is a symptom that appears at any time before the acute phase of a disorder and persists during this latter. Conversely, a residual symptom is present during the acute phase and still persists beyond it.

A series of studies conducted by Fava and colleagues investigated prodromal symptomatology of unipolar affective disorders and relationship between prodromal and residual phases. The prodromal symptomatology was found preceding the full disorder by weeks or months; generalized anxiety disorder and irritability were the most common symptoms in this phase, followed by insomnia, decreasing interest for daily activities, impaired work and fatigue (Fava et al., 1990; Fava & Kellner, 1991). A review of the literature about prodromal and residual phases of unipolar and bipolar disorders confirmed that the first one could be often detected among these patients and, similarly, residual symptoms could persist although the success of treatment conducted (Fava, 1999). The rollback phenomenon was also introduced. According to it, a relationship between prodromal and residual symptoms exists; in particular, the early stage of the disorder (prodromal symptoms) will be the last to remit: after depression acute phase, many of the symptoms experienced in the early phase will be repeated in reverse order in the residual one (Detre & Jarecki, 1971). These results were then confirmed by a new and more recent review (Fava & Tossani, 2007). The most common prodromal symptoms detected in the studies there investigated comprised feelings of anxiety and generalized anxiety disorder, irritability, sleep disturbances, fatigue, worthlessness, trouble concentrating and impaired work and interests, confirming previous findings. Therefore, rollback phenomenon was also reported: in particular, it was found that 70% of the residual symptoms were present also in the prodromal phase; this percentage increased to 90% when considering generalized anxiety disorder and irritability that resulted to be once more the most experienced symptoms.

A recent study conducted by Iacoviello and colleagues (Iacoviello, Alloy, Abramson, Choi, 2010) confirmed the main results of the previous studies providing the following outcomes:

1. prodromal phase to Major Depression Disorder was detected and several of its symptoms appeared to be common across individuals; in particular, the most often experienced symptoms in this phase were sad mood, decreased interest pleasure from activities, difficulty concentrating, hopelessness, worrying, decreased self-esteem and irritability.

2. About the relationship between prodromal and residual symptoms, 3 main hypotheses were confirmed: the profiles of the prodromal symptomatology resulted to be similar to the residual one; the order of symptoms' remission was the reverse of the order of their onset (confirmation of the rollback phenomenon); a moderate significant correlation was detected between the durations of the two phases (prodromal and residual).

3. Similarity between residual symptom profile and the subsequent prodromal symptom one among individuals who experienced more than one episode of depression was found.

As emphasized by the author of this last study, it's important avoiding the risk of an overpathologization: there are "individual differences in symptom presentation and the appearance of these symptoms will not always indicate that an episode of depression is forthcoming" (Iacoviello et al., 2010).

In this framework, it seems top-flight the need for instruments able to screen the prodromal stage of the disorder. Indeed, the authors of studies previously seen emphasized the utility of an early treatment to help thwarting the development of depressive acute phase and treating depressive symptoms more effectively (e.g., Iacoviello et al., 2010; Fava et al., 1991; Fava et al, 2007).

Several studies help us to understand how much useful brief screening tools for depression could be and what is the average of their cost-utility.

A study conducted in 2002 (Sharp & Lipsky, 2002) emphasized the importance and the utility of using screening tools in primary care settings: it resulted to be an efficient way to identify individuals who need further assessment, primary care treatment or specific assessment by a mental-health specialist. The authors also reported several guidelines with indications for the need of screening: symptoms such as history of depression, chronic pain, obesity, financial strain and social withdrawal are indicative of the need for depression screening in adults while, with children and adolescents, symptoms like antisocial behavior, diminished school performance, agitation or irritability are the main indicators. Furthermore, a brief list of the most useful screening tests in primary care settings is provided: the Pediatric Symptom Checklist (Katon & Schulberg, 1992) looked as the optimal non-specific instrument for screening depression in children population; the Center for

Epidemiologic Depression Scale (Radloff, 1977) and the Beck Depression Inventory-Primary Care (Beck, Guth, Steer, Ball, 1997) with adults; the Edinburgh Postnatal Depression Scale (Cox, Chapman, Murray, Jones, 1996) with peripartum women and the Geriatric Depression Scale (Yesavage, Brink, Rose, Lum, Huang, Adeyand, 1983) and the Cornell Scale for Depression in Dementia (Alexopoulos, Abrams, Young, Shanoian, 1988) for utilization with elderly.

These results were anticipated by Kerr (Kerr, 2001) who found that self-report screening instruments are useful in primary care settings, although scores should not be used to diagnose but as indicator of need for further assessment. In a recent study, Gilbody confirmed the importance and utility of screening programs but emphasized the need to join organizational enhancements in order to achieve effective outcomes (Gilbody, Sheldon, House, 2008).

An interesting study investigated the cost-utility of the screening for depression in primary care (Valenstein et al., 2001) and found that annual and periodic screening is expensive but one-time screening is cost-effective; furthermore, it was stated that the cost-effectiveness significantly increased if treatment becomes more efficient and effective.

III. Positive *versus* negative wording of items

One of the most commonly accepted assumption about the wording of a questionnaire items concerns the use of both positively and negatively worded stems (Anastasi, 1982; Mehrens & Lehmann, 1983; Wright & Masters, 1982). A positively worded item, also defined straight item, is a stem whose content has the same direction of the construct that want to measure: accordingly, high score on this item indicates a high level of construct. Conversely, a negatively worded item, or reverse item, has the opposite direction of the construct and a high score on this stem indicates low presence of the construct under investigation. The commonly accepted habit to balance straight and reverse items when developing a psychological test has one specific aim: the avoidance of acquiescence effect (Barnette, 2000).

Acquiescence is the tendency to agree with an item regardless of its content (Chiorri, 2011). The use of negatively worded items is thought to protect against the risk of

this bias, leading individuals to better concentrate on items' content and increasing their level of attention.

Although this is a dangerous risk, several studies queried the advantage of using reverse items in the development of a psychometric tools; in particular, the criticisms moved were of three different types as clearly summarized by Barnette (Barnette, 2000):

1. A first group of studies focused on the score reliability. Schriesheim and colleagues (Schriesheim & Hill, 1981) compared three questionnaires, the first of which was composed by all straight items, the second by all reverse items and the third by mixed items. They found that the use of mixed items led to a decreasing in internal consistency and suggested to avoid the use of reverse items in order to preserve the response accuracy. The same authors (Schriesheim, Eisenbach, Hill, 1991) detected three types of reverse items: the polar-opposite, which uses a conceptually opposite term to the straight one (e.g., "happy" and "unhappy"); the negated-regular, which simply adds a negation to the straight term; and the negated-polar-opposites, which consists of a double negation and consequently confirms the straight statement. The authors of the study found that the polar-opposites and the negated-polar-opposites types led to a decreasing of the reliability in terms of internal consistency.

Similarly, in another study, internal consistency value was found to be significantly higher when all positive stems were used (Chamberlain & Cummings, 1984).

2. A second group of researchers found that straight and reverse items refers to different latent factors. Benson (Benson, 1987) and Pilotte (Pilotte & Gable, 1990) demonstrated through factor analyses that the mixed version of a questionnaire provided different factor structures compared with a version with all straight stems and another with all reverse stems. This result was recently confirmed by Chiorri (Chiorri, Anselmo, Robusto, 2009) who found that the one-dimensional structure of a psychometric questionnaire could be weakened using mixed items in the same tool. He added that, according with this result, the correlation between positively and negatively worded items was not perfectly equal to -1.00, as expected if they were opposite each other.

3. The last group of studies identified by Barnette focused on the individuals' ability to answer reverse items. For example, Benson (Benson & Hocever, 1985) and Marsh (Marsh, 1986) found that elementary students and preadolescents had many difficulties to understand the meaning and content of negatively worded items. In particular, these individuals had difficulty discriminating between the directionally oriented items and that this difficulty is inversely correlated with the reading ability of each one: to a low reading ability corresponded high level of difficulty (Marsh, 1986). Barnette, in his study, obtained the same outcomes of these researches (Barnette, 2000) and attempted to find an alternative solution to the use of negatively worded items to avoid acquiescence bias. In order to achieve this goal, he changed the order of the responses instead of the direction of the stems. Likert response alternatives were used as follows: half of them going from "strongly disagree" to "strongly agree" and half going from "strongly agree" to "strongly disagree", while all the items were positively worded. The questionnaire showed the highest level of reliability, with a Cronbach's alpha value of 0.85, and the highest item variance. Using mixed response options and straight items could avoid individuals' confusion or difficulties and, at the same time, the risk of acquiescence, without undermining the internal consistency of a psychometric tool and its response accuracy.

2. LITERATURE REVIEW

I. Validation studies of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS)

The Warwick-Edinburgh Mental Well Being Scale (WEMWBS) is a 14-items self-report measure of mental well-being. It was developed since the importance of mental well-being has been recognised in leading to positive consequences in term of health and social costs and of preventive programmes on the community.

The WEMWBS measures both the hedonic and eudemonic perspectives of mental well-being, providing information on positive affects (cheerfulness, optimism, and relaxation), satisfying interpersonal relationships and positive functioning (personal development, competence, autonomy, self-acceptance, clear thinking, energy).

The measure refers to a period included in the two past weeks and it is possible to answer each item on a 5-point Likert scale (none of time, rarely, some of time, often, all of time). All items are positively worded and the total score ranges between 14 and 70, with a high score corresponding to a high level of mental well-being.

The start point of the development of WEMWBS was the Affectometer 2, a measure of mental well-being developed in New Zealand and validated in UK. This scale was composed by 20 statements and 20 adjectives referred to both hedonic and eudemonic mental well-being, in which positive and negative items were balanced. Despite it showed good construct validity, good discriminatory validity between different population groups and acceptable test-retest reliability it had several limitations such as a level of 0.94 of internal consistency (which could suggest redundancy of some items), a high social desirability bias and a considerable length.

The authors of the development and first validation of WEMWBS (Tennant, Hiller, Fishwick, Platt, Joseph, Weich, Parkinson, Secker & Stewart-Brown, 2007) collected data from an undergraduate and postgraduate students' sample (recruited in Warwick and Edinburgh Universities), subsequently discussing it using two focus groups in Scotland and England, and from two representative Scottish population samples.

The WEMWBS showed to have not floor or ceiling effects.

On the students' samples, the results of Factor Analysis fitted a single-factor structure and this factor has been identified in mental well-being. High Cronbach's alpha (0.89) suggests that, although a good level of internal consistency, there could be the possibility of redundancy of several items. The BID-R (Balanced Inventory of Desirable Response) was administered to assess social desirability bias: the results, in contrast to the Affectometer 2, showed that this scale was less prone to this type of bias.

Moreover, construct validity was tested comparing the WEMWBS with other scales: moderately high correlations were detected with Scale of Psychological Well-being, Satisfaction with Life Scale, Short Depression Happiness Scale, positive subscale of Positive and Negative Affect Scale and the WHO-Five Well-being Index; moderately low correlations were found with the EQ-5D thermometer (a measure of physical and emotional health) and the Emotional Intelligence Scale (a measure of the ability to accurately assess one's own and others' emotions). Finally, a correlation of 0.83 after one week confirmed good level of test-retest reliability, showing that WEMWBS scores remained robust across brief period of time. The results of the two focus groups, in which were asked to mental health service users and non-users to discuss their impression on the scale, gave good evidence of face validity. The WEMWBS resulted to be clear, unambiguous and easy to use.

The study on the population samples gave good results too. It was confirmed a single-factor structure and a high level of internal consistency (Cronbach's alpha=0.91). Significant negative correlations were also detected comparing the WEMWBS with the General Health Questionnaire (GHQ-12).

A short version of the scale (SWEMWBS) was provided by an important validation study conducted later on a sample composed by 779 participants recruited from the Scottish Health Education Population Survey (Stewart-Brown, Tennant A., Tennant R., Platt, Parkinson & Weich, 2009). Three items were deleted from the original version because showed misfit to model expectations and others were deleted because of their gender/age bias. The result was a 7-item scale, strictly one dimensional and free of bias. On the other hand, several limitations were found about the face validity of the SWEMWBS: it provided a more restricted view of the concept of mental well-being than the original version, reducing the number of items

describing hedonic well-being or affective aspects of it. Anyway, the SWEMWBS could be a good tool for monitoring level of mental well-being in populations because of its robust psychometric properties and brevity.

Good psychometric properties for WEMWBS were confirmed even by a recent validation study conducted on a population sample of teenage school students aged between 13 and 16 in England and Scotland (Clarke, Friede, Putz, Ashdown, Martin, Blake, Adi, Parkinson, Flynn, Platt & Stewart-Brown, 2011). The results reported very few missing items. High positive correlations were detected with other measures of mental well-being (World Health Organisation-5, Kidscrenn-27 scale, and Mental Health Continuum-Short Form) while negative correlations were found with the General Health Questionnaire-12 and the Strengths and Difficulties Questionnaire, both measures of mental impairment. Internal consistency of the WEMWBS was confirmed by high level of Cronbach's alpha and strong positive correlations between individual items and total scores. Only test-retest reliability resulted lower than precedent studies (0.66). In this case focus groups were also conducted in which 80 students took part and confirmed that the WEMWBS was clear and simple to use. Qualitative findings suggested that the comprehension could be a problem for people younger than 13 but this was not confirmed by quantitative findings.

An Italian version of the WEMWBS was provided (Gremigni & Stewart-Brown, 2011). Even in this study, the WEMWBS confirmed its good internal consistency (Cronbach's Alpha= 0.87) and stability (one-week test-retest=0.80) showing good reliability on a general community sample. A one-dimensional factor structure was confirmed by confirmatory factor analysis and high correlations with other mental health and mental well-being were detected. As a short scale, it resulted to be an appropriate instrument for monitoring mental well-being at a population level.

Although good psychometric properties have been confirmed until now, other studies are going on to assess possible cut-off points for this scale, its sensitivity to change over time and appropriateness to assess mental well-being among ethnic minorities in UK and across other different Countries.

II. Validation studies of the Center for Epidemiologic Studies Depression Scale (CES-D)

Lenore Sawyer Radloff of the National Institute of Mental Health Center for Epidemiologic Studies developed the CES-D in 1977 (Radloff, 1977). It is a short self-report scale created to assess the depressive symptomatology in the general population. The scale is composed of 20 items that cover the main depressive symptoms. Exploratory Factor Analyses detected 4 main factors (with an emphasis on the affective component), which were proposed to converge on a single higher-order factor measuring depression. These factors were Depressed Affect, Positive Affect, Somatic and Retarded activity, and Interpersonal problems. Responders may answer each item on a four-point Likert scale (0 = rarely or none of the time; 1 = some or little of the time; 2 = occasionally or a moderate amount of the time; 3 = most or all of the time. The total score ranges from 0 to 60.

Radloff validated the CES-D on three different samples including both subjects from the general population and psychiatric patients. The four-factor structure was confirmed; internal consistency coefficients were acceptable, ranging from 0.85 in the general population to 0.90 in the psychiatric sample. Test-retest reliability showed only moderate coefficients (0.51-0.32 for time intervals ranging between 2 weeks and 12 months); concurrent validity was good with clinical and self-report criteria, showing positive correlations with scales measuring depression or general psychopathology and negative correlations with measures of positive affect.

A standard 16 cut-off point was used to detect clinical depressive symptoms in Radloff's study (i.e., a subject with a score over 16 was considered to be "at risk" of depression or in need of treatment), but the author underlined the need of further validation studies to test this cut-off.

Weissman and colleagues conducted a valuable validation study on psychiatric males and females (aged 18-65 years), following Radloff's study (Weissman, Sholomskas, Pottenger, Prusoff, & Locke, 1977). The sample included acutely depressed patients, recovered depressed patients, drug addicts, alcoholics, and schizophrenics. In this study, the CES-D demonstrated to be a sensitive tool for detecting depressive symptoms and measuring changes in symptoms over time. The standard 16 cut-off

point showed acceptable sensitivity and specificity values, when compared with Raskin Score as a criterion.

In short, the CES-D became one of the most used instruments for the epidemiological screening of depression, included in large surveys such as the Community Mental Health Assessment Survey (Radloff & Locke, 1986) and the National Health and Nutrition Examination Survey (Eaton & Kessler, 1981).

The original version of the scale was then subjected to a revision by Eaton and colleagues in 2004. The aim was to try to overcome some of its limitations. In fact, Radloff did not base the original version on the DSM definition of Major Depressive Disorder; symptoms like anhedonia, psychomotor retardation/agitation, and suicidal ideation were not included. Finally, symptoms such as weight changes, sleeping problems, feelings of worthlessness and concentration difficulties were measured using single-item scales. The result of this revision process was a tool as reliable and valid as the original scale, but more relevant to the DSM definition of depression.

i. Cross-cultural studies

Several studies translated, cross-culturally adapted and validated the CES-D.

A study validated a Chinese-language version of the CES-D on 138 Hong Kong Chinese married couples (Cheung & Bagley, 1998). Confirmatory factor analysis identified 2 factors (Depressive symptom and Interpersonal problems), which satisfactory correlated with other measures of life satisfaction and stressful life events. Li and colleagues detected the validity and reliability of this Chinese version on 313 primary school students (Li & Hicks, 2010).

Acceptable Cronbach's alpha values and a four-factor structure were confirmed for a Spanish version of the CES-D in two studies (Soler, Pérez-Sola, Puigdemont, Pérez-Blanco, Figueres & Alvarez, 1997; Vazquez, Blanco & López, 2006) on both psychiatric and nonclinical populations. This instrument also showed acceptable sensitivity and specificity values in relation to the standard 16 cut-off point (Soler et al., 1997).

Other translations of the CES-D showed satisfactory internal consistency and validity, ability to discriminate between depressed and normal people (Fava et al., 1983), and sensitivity to variations in the intensity of depressive symptomatology

(Goncalves & Faguhla, 2004). Specifically, Italian (Fava, 1983), Portuguese (Goncalves et al., 2004), Russian (Dershem, Patsiorkovski & O'Brien, 1996), Turkish and Arabic (Spijker, Van Der Wurff, Poort, Smits, Verhoeff & Beekman, 2004) adaptations of the CES-D have been developed. Fuhrer (Fuhrer et al. 1989) developed a French version of the CES-D and found that this scale was sensible and specific at different cut-off points on the basis of gender: 17 and 23 were recommended cut-off points for males and females, respectively. The scale showed to be valid and reliable for use in hospitalized patients, outpatients, and patients consulting general medical doctors.

Furthermore, a Greek version has been developed by Fountoulakis (Fountoulakis, Iacovides, Kleanthous, Samolis, Gougoulis, Tsiptsios & Kaprinis, 2001) showing satisfactory Cronbach's alpha value (0.95) and satisfactory test-retest reliability (Pearson's r between 0.45 and 0.95 for individual items and 0.71 for the total score). Sensitivity and specificity exceed 90% at 23/24 cut-off points, and a three-factor structure emerged, including Positive affect; Irritability and interpersonal relationships; Depressed affect and somatic complains.

Campo-Arias and colleagues (Campo-Arias, Diaz Martinez, Rueda Jaimes, Cadena Afanador & Hernandez, 2007) validated the CES-D in Colombia and suggested 20 as the best cut-off point for the general population. Cronbach's alpha value was 0.87 and four factors emerged, explaining 50.3% of the total variance.

Finally, a Dutch translation of the CES-D (Haringsma, Engels, Beekman & Spinhoven, 2004) appeared to be a useful instrument for measuring depressive symptoms in the elderly (high Cronbach's alpha values: 0.80-0.90).

A recent review (Kim, Decoster, Huang & Chiriboga, 2011) analyzes racial/ethnic differences regarding the factor structure of the CES-D in the studies conducted between 1977 and 2010. The authors identify 28 studies including results from both Confirmatory Factor Analysis (CFA) and Exploratory Factor Analysis (EFA) among five racial groups: African Americans, American Indians, Asians, Whites, and Hispanics. Results of EFAs indicate that the CES-D factor structure varies significantly between groups, while CFA results replicate the Radloff's four-factor structure. Among African Americans, two additional factors (Demoralization and Distress) have been added; one factor (Alienation) has been added in the Asian

sample, and one (Preoccupation) in the White sample. Moreover, while, among Asians and Hispanics, the CES-D shows the most similar structure, among African Americans and Hispanics, the scale shows the least similar structure. When comparing individual factors between groups, only Depressed Affect (DA) and Somatic Symptoms (SS) have been uniquely defined in each group. Positive Affect (PA) is equivalent for Asians, Whites, and Hispanics, but PA for African Americans differs from the PA for Asians and Whites. The Interpersonal Problems (IP) factor is equivalent between African Americans and Asians, whereas it is not equivalent between Whites and Hispanics and other groups.

On the other hand, CFA shows that DA, SS and PA are robust across all groups and that IP is equivalent for all groups except Asians.

ii. Sensitivity and specificity of the CES-D

Several studies tried to identify the cut-off point that maximizes sensitivity and specificity values for the CES-D. Some of these (e.g., Shinar, Gross, Price, Banko, Bolduc & Robinson, 1986; Parikh, Eden, Price, & Robinson, 1988; Soler et al., 1997; Sthal et al., 2008) confirmed the acceptability of the Radloff's standard 16 cut-off point, whereas other authors (e.g., Fuhrer & Rouillon, 1989; Furuwaka, Hirai, Kitamura & Takahashi, 1997; Lewinsohn, Seeley, Roberts & Allen, 1997; Morin, Moullec, Maïano, Layet, Just & Ninot, G.2011) proposed different cut-off points on the basis of various needs.

a. Non-Clinical Population

Myers (Myers & Weissman, 1980) and Roberts (Roberts & Vernon, 1983) are among the first authors who conducted a validation of the CES-D on adult community samples. They tested the ability of the tool to screen people at risk of clinical depression symptoms. Roberts found that, with a cut-off point of 16, the CES-D is reasonably good at screening true negatives but not efficient in discriminating true positives. Using the Schedule for Affective Disorders and Schizophrenia (SADS) as research diagnostic criteria, Myers and Weissman found only modest relationships between self-reported symptoms of the CES-D (with 16 cut-off) and diagnosis of major or minor depression.

Other studies proposed different cut-off points to be used in community samples. Lewinsohn (Lewinsohn et al., 1997), using a diagnostic interview as a gold standard, found that a cut-off of 12 was the best in maximizing sensitivity and specificity on a sample of community-residing older adults. In a similar sample, a cut-off point ranging between 18 and 22 has been proposed, to achieve sensitivity of at least 80% (Dozeman et al., 2011).

To test the CES-D ability to identify cases of major depression and dysthymia (based on the DSM III R) on a sample of nine Senior High Schools in Oregon, Roberts identified a cut-off of 24 for females and 22 for males (Roberts, Lewinsohn & Seeley 1991). Although these cut-off points demonstrated good sensitivity, they performed poorly in specificity.

Several researchers carried out studies to cross-validate the CES-D in multiethnic communities.

Madianos (Madianos, Gournas & Stefanis, 1983) compared two cut-off points (i.e., 16 and 20) on the ability to screen for clinical symptoms of depression in a community sample of Greek elderly residents. Results were similar for both the cut-off points: a cut-off point of 16 led to 83.4% sensitivity and 85.9% specificity while a cut-off of 20 led to slightly higher sensitivity (85.7%) and lower specificity (84.6%). Vasquez (Vasquez et al., 2006) administered the SCID-CV and the CES-D on a sample of 554 subjects aged 18–34 years; he found that the optimal cut-off point was 26, with 90.6% sensitivity and 91.8% specificity. The Spanish version of the CES-D was also administered to Colombian samples identifying a cut-off of 23 for adolescents, with acceptable sensitivity and specificity values (i.e., 73% and 73.3%, respectively) (Camacho, Rueda-Jaimes, Latorre, Navarro-Mancilla, Escobar & Franco, 2009). A cut-off of 20 was identified for adults (sensitivity = 96% and specificity = 73%) (Campo Arias et al., 2007).

A study conducted in 2004 (Haringsma et al., 2004) on Dutch elderly community-residents who were self-referred to a prevention program for depression confirmed the criterion validity of the CES-D with a cut-off point of 25 for the screen of Major Depressive Disorder and 22 for clinically relevant Depression (with sensitivity values of 85% and 84%, respectively, but quite low specificity values of 64% and 60%).

A recent study re-evaluates the standard 16 cut-off point on a community of Chinese-American women, finding 100% sensitivity and lower specificity (i.e., 76%; Li et al., 2010).

b. Comparison of Clinical and Non-clinical Populations

Recent studies have compared clinical and non-clinical samples.

Morin and colleagues (Morin et al., 2011) analyze psychometric and screening properties of a French version of the CES-D in a subsample of 306 community adults not currently suffering from a Major Depressive Episode (MDE) or any mental disorder and in a second subsample consisted of 163 patients suffering from a MDE according to the DSM-IV and ICD-10 criteria. They compared a variety of cut-off points using depression diagnosis obtained from the MINI as a criterion. The best cut-off point for the total score was 19 (85.3% sensitivity and 85.9% specificity), but there were differences among sexes: the optimal cut-off points that were able to discriminate between clinical and nonclinical subjects were 16 for females and 20 for males.

A study assessed the utility of the CES-D in Spanish elderly samples with or without cognitive impairment (Latorre et al., 2011). In the group with cognitive impairment, the area under the ROC curve was 0.84 (95% CI: 0.77-0.89) and the cut-off point for possible depression was 13, while in the group without cognitive impairment the area was 0.90 (95% CI: 0.79-0.96) and the optimal cut-off point was 28. Furthermore, Fountoulakis and colleagues compared a sample of depressed patients with normal controls and found that a cut-off point of 23/24 was able to detect clinical depression with both sensitivity and specificity values exceeding 90% (Fountoulakis et al., 2001).

c. Clinical Population

Several studies showed the validity of the standard 16 cut-off point in various clinical populations. It resulted to be effective in screening clinical depression symptoms among patients with Breast Cancer (Margetic & Margetic, 2004), Multiple Sclerosis (Pandya, Metz & Patten, 2005), Rheumatoid Arthritis (Martens, Parker, Smarr, Hewett, Ge, Slaughter & Walker, 2006), Diabetes, and psychiatric disorders (Major Depressive Disorder, Bipolar Disorder, Adaptive Disorder with depressive mood and

other mood disorders; Soler et al., 1997). Studies conducted with stroke patients showed that the 16 cut-off point was valid (Parikh et al., 1988), but it also generated some false negatives (Shinar et al., 1986). Moreover, it gave satisfactory sensitivity but low specificity in a sample of orthopedic and neurological patients (Caracciolo & Giaquinto, 2002).

A study conducted on a sample of mothers of children with chronic disabilities evaluated the validity of the CES-D, compared to DSM-III diagnoses of major depression and generalized anxiety, using the National Institute of Mental Health Diagnostic Interview Schedule (DIS) (Breslau, 1985). This study gave an interesting result: there were not differences between these disorders, but the CES-D appeared to be as useful for detecting major depression as the GAD (using a cut-off point of 16). Therefore, these findings do not support the notion that the CES-D measures depression specifically. Another study hypothesized that the CES-D is a measure of demoralization, which could be a precursor of the development of a depressive or anxiety disorder (Roberts et al., 1989).

Finally, several studies have been carried out on clinical samples using cut-off points different from 16. Cut-off points higher than the standard one, ranging mostly from 20 to 26, have been proposed for various clinical populations.

A French validation study of the CES-D identified a cut off of 17 for males and 23 for females on a sample composed by hospitalized patients, outpatients and patients consulting general medical doctors (Fuhrer et al., 1989). In outpatient samples from community mental health centres and primary medical care centres (Schulberg, Saul, McClelland, Ganguli, Christy & Frank, 1985) the CES-D was compared with the Diagnostic Interview Schedule (DIS, DSM-III criteria). A large range of possible cut-off points has been tested. With medical patients, a good level of sensitivity was achieved with cut-off points ranging between 16 and 27, while specificity jumped from 38.6% to 70.4%. More difficult was to establish a reasonable cut-off point for psychiatric patients: even in this case the specificity was low for each cut-off point established.

A study conducted in Japan (Furukawa et al., 1997) tested the validity of the CES-D on first-visit patients to 23 psychiatric hospitals and clinics. The gold standard was the diagnosis conducted by psychiatrists on the basis of DSM III criteria. A variety of

solutions was proposed in relation to various contexts and needs. When major attention to the sensitivity was required, a cut-off point of 26 was recommended; when sensitivity and specificity were equally weighted, a cut-off point of 31 gave 74% of both the values; when specificity was more salient, a cut-off of 34 was recommended.

For other specific clinical contexts, different cut-off points have been proposed, which maximised specificity and sensitivity: 23 for HCV-Infected Injection drug users (Golub, Latka, Hagan, Havens, Hudson, Kapadia & Campbell, 2004), 24 for patients with Major Depressive Disorder, 20 for patients with any other mood disorder or Systemic Lupus Erythematosus (Julian, Gregorich, Tonner, Yazdany, Trupin, Criswell & Yelin, 2011), and 23 for patients with Systemic Sclerosis (Thombs, Hudson, Schieir, Taillefer & Baron 2008). Finally, a cut-off point lower than the standard 16 was detected for a sample of women attending family doctors (Costello, Devins & Ward, 1989): a choice of 13 resulted in zero false negatives.

III. Conclusion

The WEMWBS showed good level of validity and reliability (in terms of both stability and internal consistency); moreover, because of its shortness and good psychometric properties, it resulted to be useful in monitoring mental well-being in general population.

CES-D is a valid and reliable instrument for the screening of probable cases of depression on both general and clinical populations. In particular, it showed good reliability values in terms of internal consistency (High Cronbach's alpha value) and good content, criterion, divergent and convergent validity.

About the cut-off point used to screen people at risk of depression, the most used one is the 16 standard point but alternative cut-points have been identified among clinical and non-clinical populations and among different race/ethnic groups.

We decided to consider two cut-off points in this study: the 16 one to screen between psychologically distressed individuals and subjects at probable risk of depression, as suggested by Radloff and confirmed by several subsequent studies; the 26 cut-off for the screening of people with possible Major Depression. The use of this last more restrictive cut-off point was justified by several papers. In particular, Zich (Zich,

Atkisson & Greenfield, 1990) raised the need of a higher cut-point: he found that lower CES-D cut-off “produced too many false positives when standard (low) cut-off scores were applied. However, when stringent cut-off scores were used, results suggested that either the CES-D or BDI might assist physicians in reliably detecting depressed patients, without an overload of false positives”.

According to this statement, the study reported in this review (Vazquez et al., 2007) suggested that the 26 cut-off point showed sensitivity and specificity values exceeding 90% in a population really able to self-test because of its age (comprised between 18 and 34). Furthermore, the classification based on this cut-off have been used in a number of studies by Ensel (Ensel, 1986), Zich (Zich et al., 1990), Logsdon (Logsdon, McBride & Dean, 1994) Geisser (Geisser, Roth & Robinson, 1997).

3. THIS STUDY

I. Introduction and objectives

The mental health epidemiologic study conducted by the Mental Health Foundation showed alarming facts around mental health in UK; in particular, depression is one of the most widespread mental disorders in this population (see Introduction). In this context, methods for the screening of this disease are increasingly required.

The WEMWBS is a valid and reliable tool for assessing mental well-being at a general population level. None of the studies have investigated the ability of this tool to discriminate subjects with mental illness; nevertheless, in the last years the request of a WEMWBS cut-off point for screening depressive symptoms is getting out among Mental Health Providers and Helplines. In particular, it is guessed to be a useful questionnaire for monitoring changes in mental health over time: its positively-worded items resulted to be better accepted than the negatively worded - ones (that usually compose questionnaires for the screening of mental illness) by the general population.

The first hypothesis of this study is that the WEMWBS is a useful questionnaire for screening depression cases as well as monitoring mental well-being status. We rely to find a WEMWBS cut-off point able to discriminate between depressed and non-depressed people with high level of accuracy and sensitivity and specificity values of at least 0.8. This possibility is based on the hypothesis, to be verified, that our tool's scores are strongly negatively correlated with the CES-D scores. We suppose that higher mental well-being is associated with decreased mental impairment and depressive symptoms.

Moreover, we want to investigate if the WEMWBS is able to screen depression cases, at the cut-off point identified by this study, also in a different culture/context. The third hypothesis of this research is that the WEMWBS cut-off point identified to screen depressive cases is able to discriminate depressed and psychologically distressed individuals in an Italian sample recruited from the general population.

II. Procedure and participants

The first part of this research was conducted at the department of Mental Health of the University of Warwick, UK. For the analyses, we used existing data from a randomised controlled trial (RCT) of the Mood GYM, an internet-based cognitive behaviour therapy (CBT) programme. The RCT aimed to investigate whether such a self-delivered online CBT-based training programme can improve mental well-being of the general adult population. This was a two-arm trial, consisting of a waiting-list control arm and an intervention arm receiving access to Mood GYM, a 5 week online CBT program. In order to investigate the differences between the two treatment arms (control/Mood GYM) and the patterns of change over time, a set of scales, comprising CES-D and WEMWBS, was applied at baseline, 6-week follow-up (Posttest1) and 12 week follow-up (Posttest2).

Baseline characteristics of the sample are reported in table 1. The subjects were 3070 users of the NHS Choices website who opted into the trial advertised as a mental fitness intervention for the general population. They were a self-selected sample drawn from a general sample of internet users accessing this health portal. The most part of this sample (77.9%) was composed by females. The mean age was 41.1 (SD = 12.9). 46.7% of the participants were married while the remaining part declared to be divorced/separated, never married or widowed. 92.5% of the sample was English/Welsh/Scottish/Northern Irish/British while the others participants belong to different ethnic groups (i.e. Arab, Caribbean, Indian, Pakistani, Bangladeshi, Chinese, African and Asian). Most participants were in paid-employment or self-employed. 88.2% of the sample declared to use Internet at least once a day, 10.9% several times a week while the remaining subjects indicated a period of once a week, several times a month or less than once a month. Half of the sample (52.4%) asserted to have an excellent ability to use Internet, 37% reported a good ability while the other participants reported fair, poor, bad and not ability. Most of the participants reported a good level of general health and affirmed not be limited in day-to-day activities because of disabilities or health problems. 98.1% declared to have used Internet to find out information about health or health care; only 7.6% had used a website training programme for mental wellbeing, 7.1% had used website training

programme to treat depression or anxiety and 21.1% had experienced a Cognitive Behavioural Therapy in course of his/her life.

Table 1. Baseline characteristics of the UK sample

Variable	Sample at baseline (n = 3117)
Gender:	
Female	2421 (77.7%)
Mean (s.d.) age in years	41.10 (12.98)
Ethnicity:	
White	2777 (89.1%)
Mixed	49 (1.5%)
Asian	65 (2.1%)
Black	20 (0.6%)
Other	30 (1.0%)
Invalid/missing	176 (5.6%)
Marital status:	
Married/cohabiting	1508 (48.4%)
Divorced/separated	510 (16.4%)
Never married	1095 (35.1%)
Invalid/missing	4 (0.1%)
Employment status:	
Working	1867 (59.9%)
Student	120 (3.8%)
Retired	194 (6.2%)
Looking after home/family	327 (10.5%)
Unemployed	153 (4.9%)
Other	473 (14.4%)
Smoking:	
Daily	441(14.1%)
Occasionally	216 (6.9%)
Not at al	2455 (78.8%)
Invalid/missing	5 (0.2%)
Mean (s.d.) units of alc. last wk	39.24 (45.50)
Drug use last week:	
Yes	93 (3.0%)
Previous MH service use:	
Yes	1745 (56.0%)
Previous CBT experience	
Yes	658 (21.1%)
Previous internet-based CBT	
Yes	222 (7.1%)

The second part of this study was conducted at the University of Bologna. Data were collected by several trainees on the general population with the snowball/chain sampling method. This method establishes that subjects previously selected for the

study recruit other individuals among their acquaintances: in this way, the sample group gradually grows like a rolling snowball.

The sample was composed of 130 subjects, 52 males and 78 females, with mean age of 31.4. More than half of the sample (53.1%) affirmed to have attended 4 years of study, 55.4% was composed of students, 36.1% of employees and 8.5% of retired. The remaining part was composed of other kind of employment. The most part of the sample (60.8%) declared to live with the family of origin, 28.5% with the current family, 8.5% lived with other people while 2.3% lived alone (Table 2).

Table 2. Baseline characteristics of the Italian sample

Variable	Sample at baseline (n = 130)
Gender:	
Female	78 (60%)
Mean (s.d.) age in years	31.44 (14.98)
Cohabitation:	
Family of origin	79 (60.8%)
Current family	37 (28.5%)
Other people	11 (8.5%)
Alone	3 (2.3%)
Employment status:	
Working	48 (36.1%)
Student	72 (55.4%)
Retired	11 (8.5%)

I. Measures

WEMWBS. It is a self-report measure of mental well-being, composed of 14 items (Tennant et al., 2007). The items refer to a period included in the two past weeks. Each item is measured on a 5-point Likert scale (none of time, rarely, some of time, often, all of time). All items are positively worded and the total score ranges between 14 and 70, with a high score corresponding to a high level of mental well-being (see “Literature Review”).

WEMWBS-Italian version: like the original version, this scale measures mental well-being considering both the hedonic and eudemonic perspectives (Gremigni & Stewart-Brown, 2011). It showed good psychometric properties: high reliability in term of both internal consistency and stability over time, high correlations with other

instruments for the measuring of mental well-being and not social desirability bias were found. Factor analysis detected the existence of one main factor and suggested to delete two items in this version. The rating scale used and the scoring method are the same of the original version.

CES-D. This is a self-report scale created to assess depressive symptomatology in the general population (Radloff, 1977). Several authors also found evidence of the ability of this scale to detect generalized anxiety symptoms (e.g. Breslau, 1985). The tool is composed of 20 items, covering the main depressive symptoms (Depressed Affect, Positive Affect, Somatic and Retarded activity, and Interpersonal problems). Responders may answer each item on a four-point Likert scale (0 = rarely or none of the time; 1 = some or little of the time; 2 = occasionally or a moderate amount of the time; 3 = most or all of the time). The total score ranges from 0 to 60 with higher values representing greater psychological distress (see “Literature Review”).

PGWBI. The original version of this scale was developed by Harold Dupuy, psychologist of the National Center for Health Statistics, in 1968. It was a 68 item questionnaire created to measure mental well-being and emotional/ affective distress in the American general population. The Italian version of the scale (Grossi et al., 2002) is composed of 22 items investigating six psychological dimensions: Anxiety, Depressed Mood, Positive Well-Being, Self-Control, General Health and Vitality. Each subscale consists of 3, 4 or 5 items. The level of measurement is on 6-point Likert scale and each question refers to a period included in the four past weeks. It is possible to obtain a score for each subscale and a total score ranging between 0 and 110 with higher scores corresponding to higher mental well-being.

WHO-5. It is a self-administered one-dimensional measure of psychological well-being (Beck, 1998, 2001). It is composed of 5 positively worded items covering positive mood (good spirits, relaxation), vitality (being active and waking up fresh and rested), and general interests (being interested in things). Moreover, it showed to be a reliable measure of emotional functioning and a good screener for depression. Responders may answer each item on a 6-point Likert scale from 0 to 5; total score ranges between 0 and 25 with higher values depicting better psychological well-

being status. A score below 13 indicates poor well-being and is an indication for testing for depression under ICD-10

II. Statistical analysis

All statistical analyses were conducted with the statistical software package SPSS 19.

i. Statistical analyses on the UK sample

In the first study, descriptive statistics concerning participants' characteristics were evaluated at baseline. Frequencies and distributions of WEMWBS and CES-D scores were investigated at baseline, 6 week follow-up and 12 week follow-up.

The diagnostic performance of the WEMWBS over a range of cut-off values was analyzed by calculating Receiver Operating Characteristics (ROC) curves at baseline, post-test 1 and post-test 2. CES-D was used as gold standard at the two different cut-off points identified by a literature review as useful for a screening of depression in the general adult population: the CES-D 16 cut-off point was used to screen individuals with psychological distress and need for further assessment; the 26 cut-off point was utilized to screen subjects with probable symptoms of Major Depression (see the Literature Review section). It was specified that a smaller test result indicated more positive test: a lower score on the WEMWBS means the subject is more likely to be at risk of depression; conversely, higher score on the CES-D indicated positive (depressed) case. We obtained an ROC curve for how well each WEMWBS result reflected the CES-D diagnosis (depressed/non-depressed), an estimate of the area under the curve (AUC) and a table of the curve's coordinates (sensitivity and 1-specificity or false positive rate). The specificity and false negative rates were calculated as well. ROC curves display the relationship of sensitivity to 1-specificity: the sensitivity is the proportion of true positive cases (on the basis of CES-D diagnosis) with WEMWBS results lower than the cut-off point; 1-specificity is the proportion of true negative cases with WEMWBS results lower than the cut-off (false positive cases). The Area Under the Curve (AUC) can be considered as the probability of correct prediction and an index for determining the accuracy of the test: it represents the probability that the WEMWBS result for a randomly chosen positive case will be higher than the result for a randomly chosen negative case. The

closer is the AUC to 1, the better the predictive power of the WEMWBS: an AUC equal to 1 indicates a perfectly informative test with 100% of both sensitivity and specificity (Greiner et al., 2000). Considering that the best cut-off point is the closest to the upper corner to the left on the Cartesian coordinate system, we used the following formula (Perkins & Schisterman, 2006):

$$d = \sqrt{(1 - Se)^2 + (1 - Sp)^2}$$

where “d” is the distance from the upper corner on the left, “Se” is the sensitivity and “Sp” is the specificity value.

In order to obtain a significant cut-off point, we applied the second formula proposed by Perkins and Schisterman as well. This formula was used to calculate the Youden’s index, which evaluates the maximum vertical distance from the curve to the chance line or positive diagonal; this index is proposed by the two authors as the optimal method in the sense that it minimizes the misclassification rate. The formula follows:

$$J = Se + Sp - 1$$

Youden’s index value is comprised between 0 and 1: the closest is it to 1, the most accurate the performance of the screening test.

Finally, we run scatter plot graphs (Fig. 4) and Pearson’s r correlation between CES-D and WEMWBS total scores at baseline, post-test 1 and post-test 2 in order to verify their correlation trend over time.

ii. Statistical analyses on the Italian sample

In the second part of the study conducted on the Italian sample, frequencies and distributions of the WEMWBS, WHO-5 and PGWBI scores were evaluated.

In order to verify the ability of the WEMWBS to screen possible and probable depressed and non-depressed subjects respectively, the total sample was dichotomized in positive (depressed) and negative (non-depressed) groups on the basis of the two cut-off points detected with the previous analyses (16 cut-off point for the screening of psychological distress and need for further assessment and 26 cut-point for the screening of probable cases of Major Depression); individuals with a score under the cut-off were considered distressed/depressed while subjects with a

total score above the cut-off point were considered “non-cases”. Therefore, several two-way univariate ANOVA analyses were conducted to verify the differences between the two groups (positive and negative on the basis of the first cut-off point) in the PGWBI scores, a second one was implied to compare the two groups in the WHO-5 scores and a third one to compare the groups in the Depression subscale of the PGWBI. The same procedure was followed considering the second cut-off score identified. The “gender” was also selected as independent variable and kept under control. In this way we wanted to simultaneously test for the following null and alternative hypotheses:

1. Ho: there are not significant differences in mental well-being between positive and negative subjects as screened by the WEMWBS cut-off point previously identified;
H1: there are significant differences in mental well-being between the two groups.
2. Ho: there are not significant differences in mental well-being between males and females;
H1: there are significant differences in mental well-being between the two groups.
3. Ho: there is not a significant interaction between these two independent variables.
H1: there is a significant interaction between the two independent variables.

We used a significance level of 0.05 (95% certainty that the differences among the means of the groups are meaningful and not the result of random chance).

Before running the ANOVA, we tested the main assumptions for this analysis:

1. The distributions of errors for each dependent variable must be normally or approximately normally distributed.
2. The observations must be independent.
3. The samples must be obtained from populations of equal variances.

The non-parametric Mann-Whitney test was executed because of different size of each group considered in the ANOVA analyses. It was used to compare differences between “positive” and “negative” group in the score on the three scales

administered (PGWBI, PGWBI Depression-subscale, WHO-5) and to investigate if also in this case these differences are statistically significant.

4. RESULTS

I. Results on the UK sample

The main baseline characteristics of this study population are summarized in Table 1. The mean score of the WEMWBS was 42.08 (SD = 10.11). Skewness and kurtosis values were -.383 and .631 respectively with data slightly skewed left. The skewness and kurtosis values indicated that WEMWBS scores were normally distributed. Similar results were reported for the CES-D score; the mean score was 23.24 (SD = 13.44), with a skewness of .351 and kurtosis of -.669. The test to verify the normality of the distribution showed data skewed right but fairly normal distributed.

In Figures 1, 2 and 3, ROC curves for 16 CES-D cut-off point are reported (results at baseline, post-test 1 and post-test 2).

Figure 1. ROC curves for CES-D 16 cut-off point as gold standard. Results at baseline

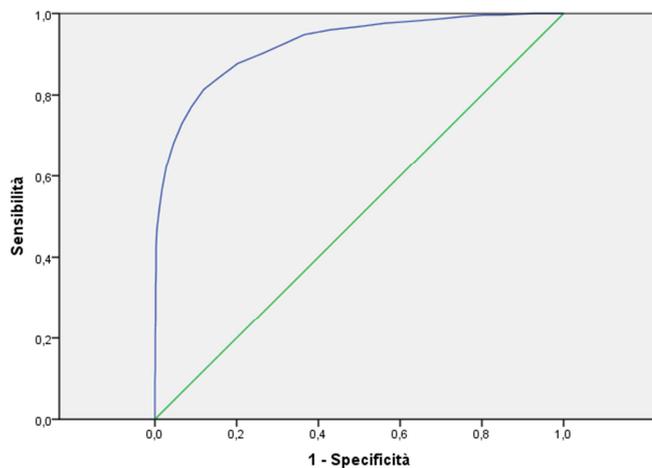


Figure 2. ROC curves for CES-D 16 cut-off point as gold standard. Results at 6-week follow-up

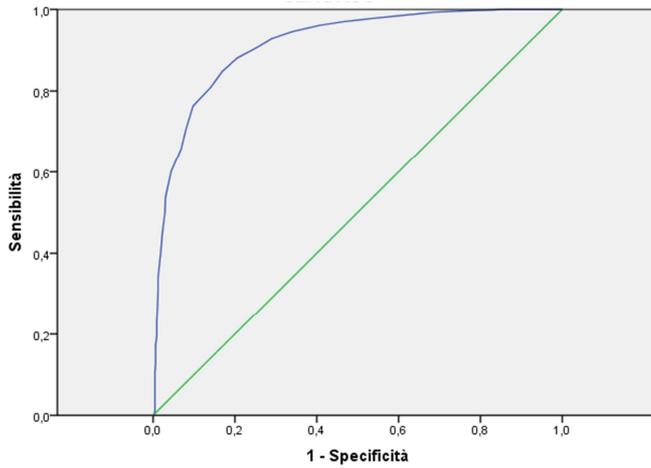
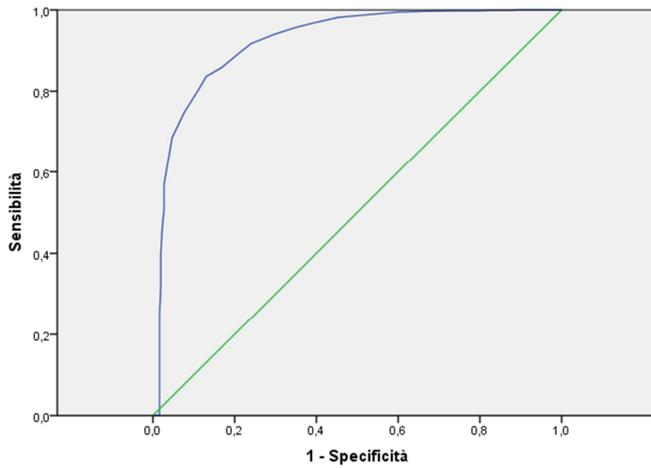


Figure 3. ROC curves for CES-D 16 cut-off point as gold standard. Results at 12-week follow-up



As we can see, it is not simple to detect the optimal cut-off point just looking at the curve. For this reason we used both the measure of the distance from the upper corner on the left and the Youden Index in order to obtain the best cut-off point in terms of screening properties. Considering a CES-D 16 cut-off point as gold standard, both the Distance value and the Youden Index suggested the same cut-off points; the 44.5 score resulted to be the optimal cut-off at baseline ($d = 0.221$; $J = 0.694$) and 12 week follow-up ($d = 0.209$; $J = 0.706$) while the 45.5 performed better at 6 week follow-up ($d = 0.227$; $J = 0.679$). According with these results, we decided

to select the 44.5 cut-off point as the optimal score with the smallest distance from the upper corner on the left and the largest vertical distance from the chance line.

Table 3. Sensitivity, specificity, AUC, false negative and false positive rates for each WEMWBS cut-off point with the smallest distance from the upper left corner of the Cartesian system and the larger Youden index. Results at baseline, 6 and 12 week follow-up

CES-D	WEMWBS CUT-OFF	AUC	SENS	SPEC	FPR	FNR
16	44.5	0.924	81.4	88.0	12.0	18.6
16	45.5	0.916	84.8	83.1	16.9	15.2
16	44.5	0.924	83.6	87.0	13.0	16.4

In Table 3 sensitivity, specificity, area under the curve, false negative and false positive rates are reported for each WEMWBS cut-off point with a CES-D 16 cut-off point as gold standard (results at baseline, post-test 1 and post-test 2). In this table we reported only the cut-points with the smallest distance from the upper corner on the left of the graph (calculated with Perkins and Schisterman's formula) and the largest Youden index.

The WEMWBS showed to be a good tool for the screening of mental distress and individuals at risk of depression with a cut-off point. At baseline, the cut-off 44.5 performed as the optimal score in screening depressive symptoms with an AUC of 0.924; according to the following Swets' criteria (Swets, 1998):

0.90-1 = excellent

0.80-0.90 = good

0.70-0.80 = fair

0.60-0.70 = poor

0.50-0.60 = fail

This result depicted an excellent ability of the WEMWBS at this time-point to screen individuals at risk of depression and need for further assessment. Sensitivity value was 81.4%, while specificity was 88%. False positive rate was 11.1% while false negative value was 18.6%.

At 6 week follow-up, 45.5 showed to be the optimal cut-off point in terms of screening properties. The AUC value was 0.916 and it depicted once more an

excellent screening property. Sensitivity and specificity values were of 84.8% and 83.1%, with false positive and false negative rates of 16.9% and 15.2% respectively. At 12 week follow-up, the 44.5 performed again as the optimal cut-off, with an AUC of 0.924. Sensitivity and specificity percentages were 83.6% and 87% while false positive and false negative rates were 13% and 16.4% respectively. ROC curves for CES-D 26 cut-off point follow (Figure 4, 5 and 6).

Figure 4. ROC curves for CES-D 26 cut-off point as gold standard. Results at baseline

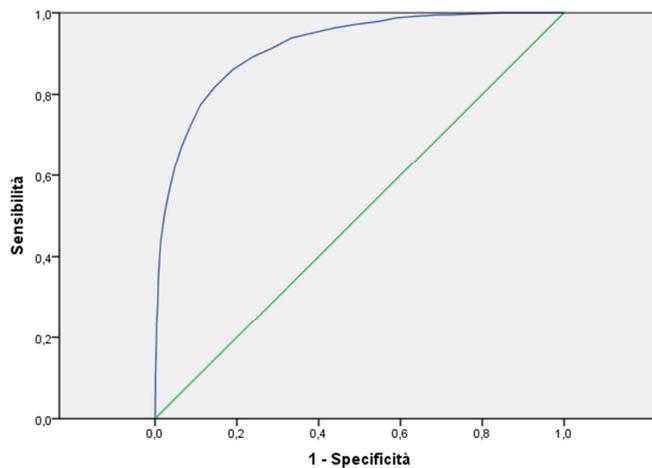


Figure 5. ROC curves for 26 CES-D cut-off point as gold standard. Results at 6-week follow-up

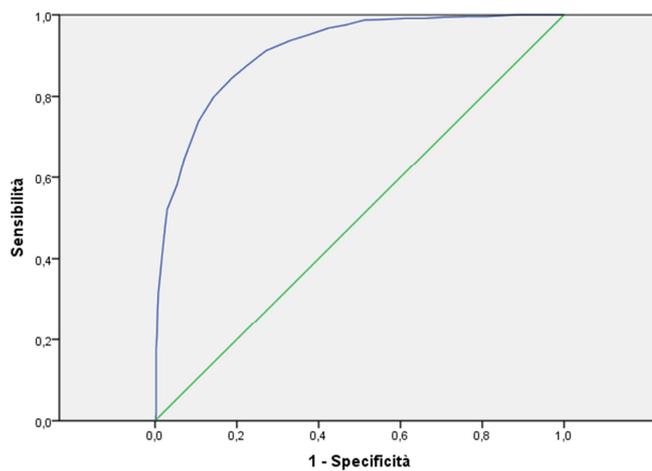
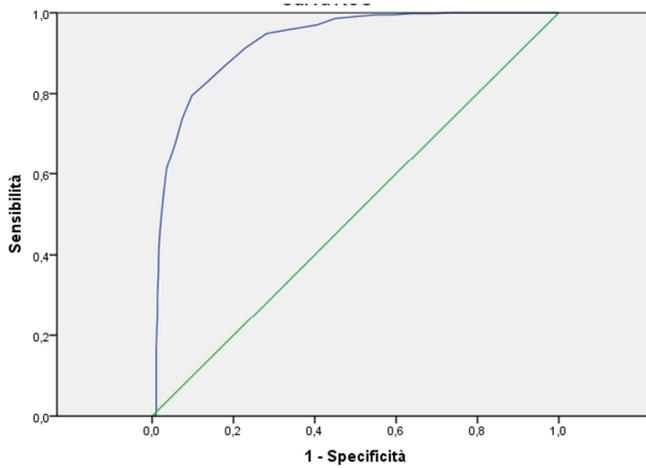


Figure 6. ROC curves for CES-D 26 cut-off point. Results at 12-week follow-up



The score 40.5 performed as the optimal cut-off point at a CES-D 26 cut-off level. In particular, at baseline the optimal score was 40.5 with both the methods with $d = 0.23$ and $J = 0.672$; at 6 week follow-up the two formulas agreed once more on a cut-point of 41.5, with $d = 0.243$ and $J = 0.657$; at 12 week follow-up the Distance value suggested the 40.5 as the best cut-off ($d = 0.218$) while the Youden index gave a result of 39.5 as the optimal value ($J = 0.697$). Nevertheless, we suggest the 40.5 as the best cut-off point considering CES-D 26 score as gold standard.

Table 4. Sensitivity, specificity, AUC, false negative and false positive rate for each WEMWBS cut-off with the smallest distance from the upper left corner of the Cartesian system and the larger Youden index. Results at baseline, 6 and 12 week follow-up

TIME	CES-D	WEMWBS CUT-OFF	AUC	SENS	SPEC	FPR	FNR
Baseline	26	40.5	0.917	81.7	85.5	14.5	18.3
Post-1	26	41.5	0.912	84.4	81.3	18.7	15.6
Post-2	26	40.5	0.928	83.0	86.3	13.7	17.0

As we can in Table 4, the AUC value exceeded once more 0.90 at all three times at a 26 CES-D cut-off level; this indicated that the WEMWBS at this level performed as a highly accurate tool in screening positive and negative cases of Major Depression. At this CES-D cut-off point as gold standard, the WEMWBS score with the best screening property is comprised between 40.5 and 41.5. At baseline, the cut-off with

the shortest distance from the upper corner on the left is the 40.5 one. At this point the WEMWBS performed highly accurately in screening symptoms of Major Depression with AUC values of .917. Sensitivity value was 81.7% while specificity was 85.5%. False negative rate was of 18.3% while false positive rate was of 14.5%. At 6 week follow-up, the best WEMWBS cut-off point resulted to be the 41.5 one, with an AUC value of 0.912, sensitivity and specificity values of 84.4% and 81.3% respectively, false-positive percentage of 18.7% and false-negative value of 15.6%. At the third time of the study, the 40.5 cut-off time showed once again to be the optimal in screening Major Depression symptoms. The AUC value was higher than the previous ones (0.928). Sensitivity and specificity values were 83.0% and 86.3% respectively while positive and negative percentages were 13.7% and 17.0%.

In Figure 7, 8 and 9, the scatter plots with the correlations between WEMWBS and CES-D total scores at baseline, 6 week follow-up and 12 week follow-up are reported. These graphs showed a high negative correlation between the questionnaires: small values of CES-D score correspond to large values of WEMWBS score.

Figure 7. Scatter-plot graph. Correlation between WEMWBS and CES-D scores at baseline

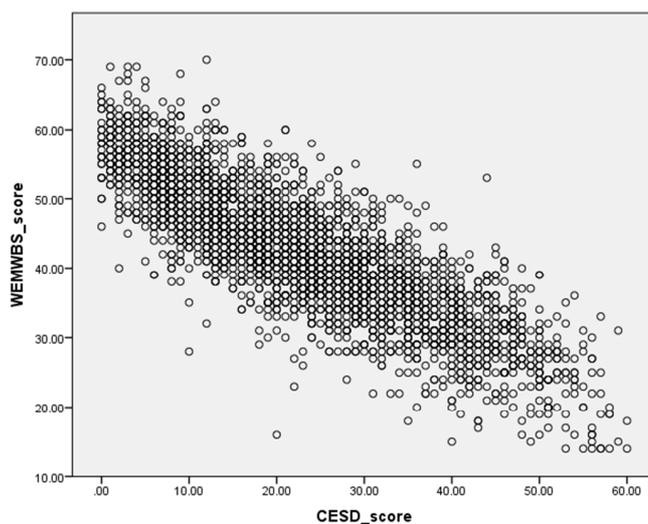


Figure 8. Scatter-plot graph. Correlation between CES-D and WEMWBS score at 6-week follow-up

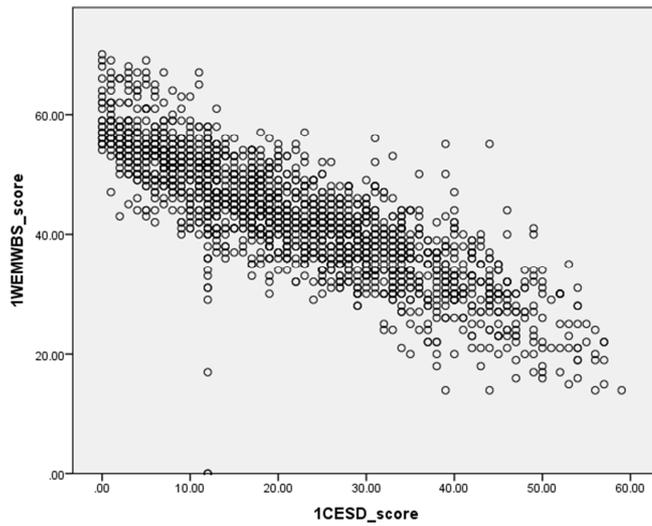
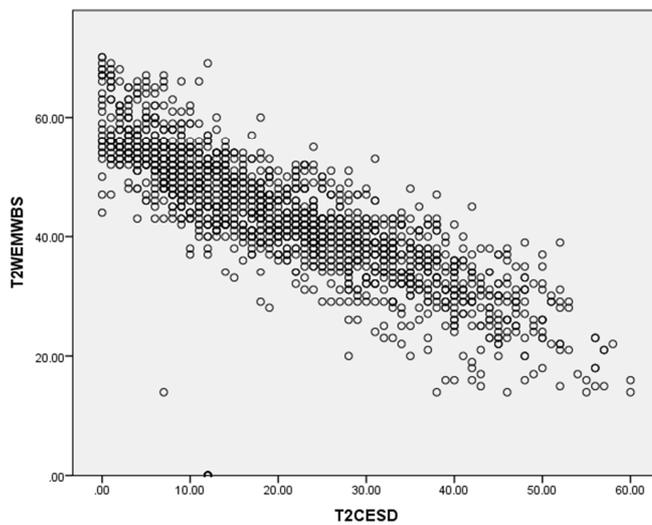


Figure 9. Scatter-plot graph. Correlation between CES-D and WEMWBS scores at 12-week follow-up



Pearson's r coefficients, calculated to investigate the graphs' results, confirmed these outcomes. Considering that r coefficient is comprised between 0 and 1 and that:

- 0 < r < 0.3: poor correlation;
- 0.3 < r < 0.7: mild correlation;
- r < 0.7: strong correlation.

The correlation coefficients can be classified as strong at each time: -.842 at baseline, -.827 at post-test 1 and -.800 at post-test 2. These correlations were significant, with a p-value smaller than 0.01 (considering the test $H_0: r=0$ vs. $r \neq 0$).

II. Results on the Italian sample

In the Italian sample the PGWBI mean score was 77.8 (SD=12.2), with a minimum score of 47 and a maximum of 104. Skewness value was -0.491 while kurtosis was -0.151, these values suggests a fairly normal distributions of the scores of this tool.

The WHO-5 mean score was 14.5 (SD=4.61), with minimum and maximum scores of 3 and 24 respectively; considering that a score below 13 is commonly considered as indicative of poor well-being and is an indication for testing for depression under ICD-10 criteria; our mean score indicates subjects not at risk of depression. Skewness and kurtosis values were -0.376 and -0.696 respectively; these values are indicative of a fairly normal distribution of the scores.

The WEMWBS mean score was 49.3 (SD=6.98) and is consistent with previous studies which found a mean score of 50 on the general population (Tennant et al., 2007). Minimum score was 20 while maximum was 68. Skewness and kurtosis values were -0.540 and 1.945, showing a fairly normal distribution of the results.

After the dichotomization of the WEMWBS score on the basis of 40.5 cut-off point, 14 individuals showed scores under the 40.5 cut-off point: 10.8% of the sample seemed to be at risk of Major Depression while 89.2% resulted not to be depressed, according to this classification. About the cut-off 44.5, 28 individuals resulted “positive”: 21.5% of the sample showed psychological distress and need further investigation of mental health status.

i. ANOVA and Mann Whitney analyses’ results for 44.5 cut-off point

Before running ANOVA, all the main assumptions were tested. The distributions of errors for each dependent variable were approximately normal and the observations were independent from each other. The first univariate two-way ANOVA was conducted to investigate the differences between “positive” and “negative” individuals (on the basis of 44.5 cut-off point) and males and females in the level of psychological distress as measured by the total score of the PGWBI.

Levene's test, which tested the null hypothesis that the population variances are equal (assumption of the homogeneity of variance with a critical p value of 0.05), showed non-significant results ($p = 0.96$): this assumption was not violated as well. The interaction effect between gender and WEMWBS screening was not statistically significant with $F = 0.174$ ($p = 0.677$). There wasn't a significant main effect for gender ($F = 2.434$; $p = 0.121$) while the main effect for WEMWBS screening was statistically significant with $F = 33.164$ and $p \leq 0.001$. Moreover, the effect size was considered large with partial eta squared of 0.21 on the basis of the following Cohen's criteria:

0.01 = small effect

0.059 = medium effect

0.138 = large effect

In this case, 21% of the change in the dependent variable (PGWBI) can be accounted for by the WEMWBS screening.

The results of Mann Whitney test confirmed these outcomes: there was a statistically significant difference between "positive" and "negative" group's median PGWBI score ($U = 431.5$, p). In Table 5 the results are summarized.

Table 5. Comparison between "positive" and "negative" subjects in PGWBI total score. Results of ANOVA and Mann Whitney's test

GROUP	MEAN (SD)	F	p
Score > 44.5	80.98 (10.72)	33.164	0.001
Score ≤ 44.5	66 (10.74)		
GROUP	MEAN RANK	U	
Score > 44.5	74.27	431.5	0.001
Score ≤ 44.5	29.98		

The second ANOVA was conducted to investigate the differences between "positive" and "negative" individuals and males and females in the level of depression as measured by the Depression subscale of the PGWBI.

Levene’s test showed non-significant results ($p = 0.052$) so the assumption of the homogeneity of variance was not violated. The interaction effect between gender and WEMWBS screening was not statistically significant with $F = 0.061$ ($p = 0.805$). There wasn’t a significant main effect for gender ($F = 3.644$; $p = 0.059$) while the main effect for WEMWBS screening was statistically significant with $F = 25.331$ and $p \leq 0.001$. The effect size was large with partial eta squared of 0.167.

The Mann Whitney test confirmed these results with a statistically significant difference between “positive” and “negative” group's median PGWBI Depression-subscale score ($U = 501.5$, $p \leq 0.001$). In Table 6 the results are summarized.

Table 6. Comparison between "positive" and "negative" subjects in PGWBI Depression-subscale score. Results of ANOVA and Mann Whitney's test

GROUP	MEAN (SD)	F	<i>p</i>
Score > 44.5	1.34 (0.614)	25.331	0.001
Score ≤ 44.5	2.14 (0.618)		
GROUP	MEAN RANK	U	
Score > 44.5	56.42	501.5	0.001
Score ≤ 44.5	98.59		

The third ANOVA was conducted to investigate the differences between “positive” and “negative” individuals and males and females in the level of emotional functioning and depression as measured by the WHO-5 total score.

Levene’s test was non-significant ($p = 0.783$). The interaction effect between gender and WEMWBS screening was not statistically significant with $F = 0.002$ ($p = 0.966$). There wasn’t a significant main effect for gender ($F = 2.974$; $p = 0.087$) while the main effect for WEMWBS screening was statistically significant with $F = 19.173$ and $p \leq 0.001$. In this case, the effect size was medium with partial eta squared of 0.132.

The Mann Whitney test confirmed the statistically significant difference between “positive” and “negative” group's median WHO-5 score ($U = 591.5$, $p \leq 0.001$). In Table 7 the results are summarized.

Table 7. Comparison between "positive" and "negative" subjects in WHO-5 score. Results of ANOVA and Mann Whitney's test

GROUP	MEAN (SD)	F	<i>p</i>
Score > 44.5	15.5 (4.33)	19.173	0.001
Score ≤ 44.5	10.82 (3.68)		
GROUP	MEAN RANK	U	
Score > 44.5	73.70	591.5	0.001
Score ≤ 44.5	35.63		

ii. ANOVA and Mann Whitney analyses' results for 40.5 cut-off point

The ANOVA conducted to test the differences between “positive” and “negative” individuals (on the basis of 40.5 cut-off point) and males and females in the PGWBI showed non-significant Levene’s test ($p = 0.875$); the interaction effect between gender and WEMWBS screening was not statistically significant as well with $F = 0.019$ ($p = 0.889$). There wasn’t a significant main effect for gender ($F = 2.454$; $p = 0.120$) while the main effect for WEMWBS screening was statistically significant with $F = 23.099$ and $p \leq 0.001$. The effect size was large with partial eta squared of 0.132.

The Mann Whitney test confirmed these results with $U = 218.5$ ($p \leq 0.001$). In Table 8 the results are summarized.

Table 8. Comparison between "positive" and "negative" subjects in PGWBI total score. Results of ANOVA and Mann Whitney test

GROUP	MEAN (SD)	F	<i>p</i>
Score > 40.5	79.55 (11.1)	23.099	0.001
Score ≤ 40.5	62.61 (11.3)		
GROUP	MEAN RANK	U	
Score > 40.5	69.62	218.5	0.001
Score ≤ 40.5	23.81		

The second ANOVA conducted to investigate the differences between “positive” and “negative” individuals and males and females in the level of depression as measured by the PGWBI depression-subscale. Levene’s test was non-significant ($p = 0.447$). The interaction effect between gender and WEMWBS screening was not statistically significant with $F = 0.005$ ($p = 0.942$). There wasn’t a significant main effect for gender ($F = 2.38$; $p = 0.125$) while the main effect for WEMWBS screening was statistically significant with $F = 20.025$ and $p \leq 0.001$. The effect size was medium with partial eta squared of 0.137. The Mann Whitney test confirmed the statistically significant difference between “positive” and “negative” group's median WHO-5 score ($U = 225.5$, $p \leq 0.001$). In Table 9 the results are summarized.

Table 9. Comparison between "positive" and "negative" subjects in PGWBI Depression-subscale score. Results of ANOVA and Mann Whitney's test

GROUP	MEAN (SD)	F	<i>p</i>
Score > 40.5	1.42 (0.64)	20.025	0.001
Score ≤ 40.5	2.3 (0.58)		
GROUP	MEAN RANK	U	
Score > 40.5	60.44	225.5	0.001
Score ≤ 40.5	107.39		

The last ANOVA was conducted with WHO-5 as independent variable. This analysis showed non-significant Levene’s test ($p = 0.977$); the interaction effect between gender and WEMWBS screening was not statistically significant with $F = 0.010$ ($p = 0.920$). There wasn’t a significant main effect for gender ($F = 2.612$; $p = 0.109$) while the main effect for WEMWBS screening was statistically significant with $F = 10.499$ and $p = 0.002$. The effect size was medium with partial eta squared of 0.077. The Mann Whitney test confirmed these results with $U = 367$ ($p \leq 0.001$). In Table 10 the results are summarized.

Table 10. Comparison between "positive" and "negative" subjects in WHO-5 score. Results of ANOVA and Mann Whitney's test

GROUP	MEAN (SD)	F	<i>p</i>
Score > 40.5	15.01 (4.41)	10.499	0.001
Score ≤ 40.5	10.35 (4.25)		
GROUP	MEAN RANK	U	
Score > 40.5	69.34	367	0.001
Score ≤ 40.5	33.61		

5. DISCUSSION

We established three main goals to achieve in this research.

The first aim was to investigate whether the WEMWBS was a useful questionnaire not only for the measurement of mental well-being, but also for the screening of mental distress and, in particular, depression. We proposed one or more cut-off points, which could be utilized in clinical and research settings for a first screening of individuals at risk for depression. We tested this hypothesis on an UK sample composed of 3070 users of the NHS Choices website. The CES-D was used as gold standard. Based on a literature review, we selected the 16 cut-off point as score for the screening of people at risk of depression, and in need for further assessment “normal” individuals, while the 26 cut-off was chosen to detect individuals with possible Major Depression. The performance of this tool in detecting possible/probable depressed people was compared with the WEMWBS one.

The second hypothesis of this study was the existence of a strong negative correlation between the scores of these two tools, the WEMWBS and the CES-D. We wanted to investigate whether to a higher score on the CES-D corresponded lower score on the WEMWBS and how much strong this correlation was.

The last aim of our research was to test the results obtained on a culturally different sample. In order to fulfill this goal, we used the cut-off points identified in the UK study on an Italian sample composed of people recruited from the general population. Then we investigated whether “positive” and “negative” individuals, assessed through the PGWBI and the WHO-5 scales, as screened by the WEMWBS cut-off points established with the UK sample, differed in the level of psychological distress and well-being.

In order to verify the first hypothesis of this study, we calculated the ROC curves comparing the screening properties of each score of the WEMWBS with the diagnosis obtained with the 16 and then the 26 CES-D cut-off points. In this way, we wanted to detect a WEMWBS cut-off for the screening of psychologically distressed people and individuals at risk of depression (corresponding to the diagnosis done by

the 16 CES-D cut-off point) and one cut-off for the screening of possible cases of Major Depression (corresponding to the screening obtained with the 26 cut-point).

To test the second hypothesis we calculated Pearson's r coefficient and depicted this correlation in a scatter-plot graph.

In order to investigate the last goal of the study, we divided the total Italian sample in two groups on the basis of score obtained at the WEMWBS: one "positive" group, if the total score was lower or equal than the cut-off points obtained through the previous analysis, and one "negative" if the score was higher. Then we executed ANOVA analyses to compare the two groups thus obtained in the score of the PGWBI, the Depression-subscale of this questionnaire and the WHO-5.

Results obtained with these analyses confirmed all the initial hypotheses.

The first part of this study, conducted on an UK sample, provided results in agreement with our initial hypotheses. About the selection of the WEMWBS cut-off point, the analyses conducted on this sample led us to choose a score of 44.5 corresponding to the 16 CES-D score and the 40.5 corresponding to the 26 one. The two methods used to evaluate the optimal cut-off (the Youden index and the Distance measure) agreed in the selection of these two scores in the three different times (baseline, 6 week and 6 week follow-up). ROC curves obtained showed good parameters: the AUC was at least 0.9 at the three times, suggesting excellent screening properties of WEMWBS in detecting depression cases as well as in measuring mental well-being status. This questionnaire also demonstrated to screen between depressed and non-depressed individuals with values of sensitivity and specificity of at least 80%: this means that it is able to detect true negative and true positive cases restraining the number of people incorrectly identified as positive (depressed) or normal (non-depressed). In particular, false positive and false negative rates provided using 44.5 and 40.5 cut-off points ranged between a minimum of 12% and a maximum of 18%.

Some more words should be said about this issue. In this study we reported also WEMWBS cut-off points that showed the optimal values in term of distance from the upper corner on the left of the Cartesian coordinate system obtained calculating ROC curves and in term of vertical distance from the chance line of the same graph. Otherwise, the selection of the best cut-off point should be a balance between

sensitivity and specificity on the basis of clinical/research needs. For example, we could need sensitivity higher than specificity if our priority is depression cases finding and minimizing missed positive cases; this could be the case of depressed subjects at risk of suicide: losing positive cases would mean obtaining very serious outcomes such as an increase in the number of suicides. In this framework we could choose a cut-off of 45.5/46.5 or higher. On the other hand, it could be required a minimum of false positive; this case could occur, for example, if further assessment of positive tested people requires expensive and invasive procedures. If we want to maximize the ability of the WEMWBS to detect the true negative cases, we could choose lower WEMWBS cut-off points (39.5/38.5 or lower).

The second important outcome of this study is the high correlation detected between the WEMWBS and the CES-D scores and the significant Pearson's r value (exceeding 0.8) at all the three times (baseline, post-test 1 and post-test 2). We found that lower WEMWBS scores were associated with increased mental impairment and increased depressive symptoms scores. Thus, a score below the WEMWBS 44.5 cut-off point reflects subjects at risk of depression while a score higher than the cut-off point depicts people not at risk of mental impairment. It is the same for the 40.5 cut-off-point: a score under this cut-point indicates possibly depressed individuals while a score above this point reflects non-clinical and non-depressed people. The strong Pearson's correlation coefficient means that the WEMWBS and the CES-D move "hand-in-hand". This conclusion is straightforward looking at the scatterplot graphs: the dispersion of the cases is nearly linear.

Interesting results has been obtained with the analyses conducted on the Italian sample. This research represents the first attempt to detect a clinical cut-off point for the WEMWBS; no other previous study tried to do this. It was important to test the first results obtained on the UK sample on another sample recruited from the general population.

The Italian sample size was not so large as the UK sample, and this may constitute a limitation of the present study. Nevertheless, results were in agreement with our initial hypotheses. Using cut-off point of 44.5, the WEMWBS showed to discriminate, in a statistically significant way, between psychologically distressed and non-distressed individuals. The ANOVA conducted showed significant

differences between “positive” and “negative” subjects in the level of mental distress and depression as measured by the PGWBI, the Depression subscale of this questionnaire and the WHO-5. Moreover, there was not significant influence of “gender” on these results. The same outcomes were obtained with the cut-off 40.5.

Taken together, these results confirmed the conclusions obtained on the UK sample. Therefore, it seems possible to use the WEMWBS not only to measure the level of psychological well-being, but also to discriminate between probable/possible depressed and “normal” individuals.

Although other tests are needed in order to generalize the results, this study constitutes a good start point for a larger use of an instrument that resulted to be well accepted by general populations and primary care patients because of its positively worded items.

Several limitations could be found in the present study.

First, the choice of the CES-D cut-off score used was tricky. A debate is ongoing on which is the optimal cut-off point for the screening of depression in the general population as in clinical samples: the original study suggested a CES-D cut-off score of 16 for the screening of depressive cases (Radloff, 1976). Additional studies conducted among the general population provided divergent cut-off scores ranging from 12 to 26 (e.g. Lewinsohn et al., 1997; Dozeman et al., 2010). Recently, cross-cultural studies found divergent results as well (e.g. Camacho et al., 2009; Campoarias et al., 2007). The great amount of proposals reported in the Literature Review chapter of this report is symptomatic of the difficulty to find an agreement to settle this issue. Despite this, it is known that to find the optimal cut-off point and make a good choice is necessary to look at the prevalence of the disorder in the population we are analyzing. In fact, this parameter affect the test performance, in particular its Positive Predictive Value, that decreases when the prevalence of the disease is low and increases if the prevalence is high in a specific population (Grimes & Schulz, 2002). Accordingly, it should be considered that the cut-off point of a screening test must be adjusted for any population. This is what we tried to do, collecting the studies which proposed cut-points used on populations similar to ours and looking for the scores mostly used in the clinical/research practice.

In addition, a criticism could be moved against the need to dichotomize psychological status and the emphasis placed on the cut-off points for the screening of a mental disorder. Moreover, it could be argued that the dichotomization always leads to a loss of (statistical) information. Although the pertinence and in part truthfulness of these possible comments, the medical world insists on dichotomizing continuous scores to facilitate the diagnostic process. In this instance, we are looking at whether it is possible to use the WEMWBS as a screening tool in the same way as the CES-D is used for the same purpose. By identifying a level of WEMWBS scores that is equivalent to a CES-D cut-off point for depression we can say that this instrument is usable to screen for depression (at least as well as any other instrument around). People above the cut-off score may be regarded as characterized by a reasonable degree of psychological well-being, while those below the cut-off can be seen potentially depressed. As we said before, many users are asking for positively worded tools to be used for the screening of psychological distress we made an attempt to respond to this request by offering two consistent cut-off points for the WEMWBS that make this well-validated questionnaire a screening tool for depression.

We suggested those cut-off points that maximized sensitivity and specificity and found the best compromise between these two parameters. It is evident that the choice of the optimal score is guided by the clinical or research's needs of the psychologists, psychiatrists, doctors and social workers who are using this questionnaire. Therefore, a high sensitivity can be preferred to a high specificity or vice versa, depending on the needs of the researcher or clinician.

6. CONCLUSIONS

The WEMWBS is a self-report measure of mental well-being at a population level. This is the first study that validates the ability of the WEMWBS to discriminate between normal persons and people at risk of Major Depressive Disorder.

This study rose from a specific request advanced by several helplines, primary care doctors and other social workers: they required a WEMWBS cut-off point for the screening of depression since this questionnaire was considered highly acceptable by the most part of patients/individuals because of its positively worded items. This last issue has been discussed in the literature. Several studies emphasized the disadvantages of using negatively worded items as they were found to lead to a decline in the tool's internal consistency (e.g., Barnette, 2000; Schriesheim et al., 1991; Chiorri et al., 2009). Furthermore, it was found that reverse and straight items have different factor structures and that the utilization of mixed stems could compromise the one-dimensionality of a questionnaire (e.g., Benson, 1987; Chiorri et al., 2009). Negatively worded items also provided the onset of difficulties in specific populations, such as children and preadolescents: this latter resulted impaired in the interpretation of these items' content and direction (Benson et al., 1985; Marsh, 1986). For these reasons, the utilization of questionnaires composed of all straight items is encouraged: it avoids a loss in internal consistency and ensures a full comprehension of the stems' content. About the risk of acquiescence, against which the reverse items are usually used, Barnette suggested to convert the response alternatives instead of the items direction: this method led to the highest internal consistency and items variance (Barnette, 2000).

Taken together, results of this study confirmed the initial hypotheses formulated. This is just the first study that tries to investigate the properties of the WEMWBS to screen depressed people. So, further confirmations are necessary to state that this questionnaire is actually useful not only for the measurement of mental well-being, but also for depression screening purposes. We expected that other researches will be conducted on both UK and Italian populations; in particular, it could be useful to develop an Italian study in which the WEMWBS score is compared with the CES-D

score, in order to replicate the procedure used on the UK sample and provide further confirmations of our results.

Another interesting orientation for future studies could concern the investigation of the sensibility change of the WEMWBS over time.

Furthermore, it could be analyzed the performance of the WEMWBS on clinical populations, considering once more the CES-D as gold standard (selecting the cut-off point adjusted for those populations), or another instrument ordinarily used for the screening of depression in the clinical practice: as reported by Sharp (Sharp et al., 2002), the Beck Depression Inventory-Primary Care is one of the most utilized tool in primary care setting for the screening of people at risk of depression. Comparing its performance with that of the WEMWBS may give further evidence of the screening properties of the WEMWBS.

REFERENCES

- Alexopoulos, G.S., Abrams, R.C., Young, R.C., & Shanoian, C.A. (1988). Cornell Scale for Depression in Dementia. *Biology Psychiatry*, *23*, 271–84.
- American Psychiatric Association (2000). *DSM-IV-TR Diagnostic and Statistical Manual of mental disorders*. (4th ed., text rev.). Washington, DC: Author.
- Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan.
- Aukst-Margetic, B., & Margetic, B. (2004). Usage of diagnostic questionnaire for depressive patients versus center for epidemiologic studies of depression scale in breast cancer patients. *Psychiatria Danubina*, *16*(4), 315-321.
- Barnette, J.J. (2000). Effects of stem and Likert response option reversals on survey internal consistency: if you feel the need, there is a better alternative to using those negatively worded stems. *Educational and Psychological Measurement*, *60*(3), 361-370.
- Bech, P. (1998). *Quality of life in the psychiatric patient*. London: Mosby-Wolfe.
- Bech, P. (2001). Male depression: stress and aggression as pathways to major depression. In: A. Dawson, & A. Tylee, “Depression: Social and Economic Timebomb”. London: BMJ Books, 63-66.
- Beck, A.T., Guth, D., Steer, R.A., & Ball, R. (1997). Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behaviour Research and Therapy*, *35*, 785–91.
- Benson, J. (1987). Detecting item bias in affective scales. *Educational and Psychological Measurement*, *47*, 55-67.

- Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22*, 231-240.
- Bradburn, N.M. (1969). *The structure of psychological well-being*. Chicago: Aldine.
- Breslau, N. (1985). Depressive symptoms, major depression, and generalized anxiety: a comparison of self-reports on CES-D and results from diagnostic interviews. *Psychiatry Research, 15*(3), 219-229.
- Camacho, P.A., Rueda-Jaimes, G.E., Latorre, J.F., Navarro-Mancilla, A.A., Escobar, M., & Franco, J.A. (2009). Validity and reliability of the Center for Epidemiologic Studies-Depression scale in Colombian adolescent students. *Biomedica revista del Instituto Nacional de Salud, 29*(2), 260-269.
- Campo Arias, A., Diaz Martinez, L.A., Rueda Jaimes, G.E., Cadena Afanador, L.D. P., & Hernandez, N. L. (2007). Psychometric properties of the CES-D Scale among colombian adults from the general population. *Revista Colombiane de Psiquiatria 4*, 664-674.
- Caracciolo, B., & Giaquinto, S. (2002). Criterion validity of the center for epidemiological studies depression (CES-D) scale in a sample of rehabilitation inpatients. *Journal of Rehabilitation Medicine, 34*(5), 221-225.
- Kahneman, D. (1999). Objective happiness. See Kahneman, & Schkade, D. (1999). Predicting the well-being effect of new circumstances: Changes are proxies for states. Working Paper: Princeton University.
- Katon, W., & Schulberg, H. (1992). Epidemiology of depression in primary care. *General Hospital Psychiatry, 14*, 237-47.
- Chamberlain, V.M., & Cummings, M. N. (1984). Development of an instructor/course evaluation instrument. *College Student Journal, 18*, 246-250.

- Cheung, C.K., & Bagley, C. (1998). Validating an American scale in Hong Kong: the Center for Epidemiological Studies of Depression Scale (CESD). *The Journal of Psychology, 132*(2), 169-186.
- Cheung, S.T., & Chang, A.C.M. (2005). Measuring psychological well-being in the Chinese. *Personality and Individual Differences, 38*(6), 1307-1316.
- Chiorri, C. (2011). Teoria e tecnica psicometrica. Costruire un test psicologico. Milano: McGraw-Hill.
- Chiorri, C., Anselmi, P., & Robusto, E. (2009). Reverse items are not opposites of straightforward items. In U. Savardi (Ed.). *The Perception and Cognition of Contraries* (pp. 295-328). Milano: Mc-Graw Hill.
- Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., Adi, Y., Parkinson, J., Flynn, P., Platt, S., & Stewart-Brown, S. (2011). Warwick-Edinburgh Mental Well-being Scale (WEMWBS): mixed methods assessment of validity and reliability in teenage school students in England and Scotland. *BMC Public Health, 11*(1), doi:10.1186/1471-2458-11-487
- Clarke, P.J., Marshall, V.W., Ryff, C.D., & Wheaton, B. (2001). Measuring psychological well-being in the Canadian Study of Health and Aging. *International Psychogeriatrics, 13*(1), 79-90.
- Costello, C.G., Devins, G.M., & Ward, K.W. (1988). The prevalence of fears, phobias and anxiety disorders and their relationship with depression in women attending family physicians. *Behaviour Research and Therapy, 26*(4), 311-320.
- Cowen, E.L. (1991). In pursuit of wellness. *American Psychology, 46*, 404-8.
- Cox, J.L., Chapman, G., Murray, D., & Jones, P. (1996). Validation of the Edinburgh Postnatal Depression Scale (EPDS) in non-postnatal women. *Journal of Affective Disorders, 39*, 185-189.
- Deci, E.L. (1975). *Intrinsic Motivation*. New York: Plenum.

- Dershem, L.D., Patsiorkovski, V.V., & O'Brien, J. (1996). The use of the CES-D for measuring symptoms of depression in three rural russian villages. *Social Indicators Research*, 39(1), 89-108.
- Detre, P., & Jarecki, H. (1971). *Modern Psychiatric Treatment*. Philadelphia: Lippincott.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542-75.
- Diener, E., Gohm, C.L., Suh, E., & Oishi, S. (2000). Similarity of the relations between marital status and subjective well-being across cultures. *Journal of Cross-Cultural Psychology*, 31, 419-436.
- Diener, E., & Larsen, R.J. (1984). Temporal stability and cross-situational consistency of affective, behavioral, and cognitive responses. *Journal of Personality and Social Psychology*, 47, 580-592.
- Dozeman, E., van Schaik, D.J.F., van Marwijk, H.W.J., Stek, M.L., van der Horst, H.E., & Beekman, A.T.F. (2010). The center for epidemiological studies depression scale (CES-D) is an adequate screening instrument for depressive and anxiety disorders in a very old population living in residential homes. *International Journal of Geriatric Psychiatry*, 26(3), 239-246.
- Eaton, W., & Kessler, L.G. (1981). Rates of symptoms of depression in a national sample. *American Journal of Epidemiology*, 114(4), 528-538.
- Ensel, W.M., Lin, N., & Dean, A. (1986). *Measuring depression: The CES-D Scale. Social Support, Life Events and Depression*. New York: Academic Press.
- Fava, G.A. (1983). Assessing depressive symptoms across cultures: Italian validation of the CES-D self-rating scale. *Journal of Clinical Psychology*, 39(2), 249-251.
- Fava, G.A. (1999). Subclinical symptoms in mood disorders: pathophysiological and therapeutic implication. *Psychological Medicine*, 29, 47-61.

- Fava, G.A., Grandi, S., Canestrari, R., & Molnar, G. (1990). Prodromal symptoms in primary major depressive disorder. *Journal of Affective Disorders, 19*, 149–152.
- Fava, G.A., Grandi, S., Zilezny, M., Canestrari, R., & Morphy, M. (1994). Cognitive behavioral treatment of residual symptoms in primary major depressive disorder. *American Journal of Psychiatry, 151*, 1295–1299.
- Fava, G.A., & Kellner, R. (1991). Prodromal symptoms in affective disorders. *American Journal of Psychiatry, 148*, 823–830.
- Fava, G.A., & Ruini, C. (2003). Development and characteristics of a well-being enhancing psychotherapeutic strategy: well-being therapy. *Journal of Behavior Therapy and Experimental Psychiatry 34*, 45–63.
- Fava, G.A., & Tossani, E. (2007). Prodromal stage of major depression. *Early Intervention in Psychiatry, 1*, 9–18.
- Fechner-Bates, S., Coyne, J.C., & Schwenk, T.L. (1994). The relationship of self-reported distress to depressive disorders and other psychopathology. *Journal of Consulting and Clinical Psychology, 62*(3), 550-559.
- Fountoulakis, K., Iacovides, A., Kleanthous, S., Samolis, S., Gougoulas, K., Tsiptsios, I., Kaprinis, G., et al. (2003). Reliability, validity and psychometric properties of the Greek translation of the Major Depression Inventory. *BMC Psychiatry, 3*, doi:10.1186/1471-244X-3-2
- Fuhrer, R., & Rouillon, F. (1989). La version française de l'échelle CES-D (Center for Epidemiologic Studies-Depression Scale). Description et traduction de l'échelle d'auto-évaluation. *Psychiatry Psychobiology, 4*, 163-166.
- Furukawa, T., Hirai, T., Kitamura, T., & Takahashi, K. (1997). Application of the Center for Epidemiologic Studies Depression Scale among first-visit psychiatric patients: a new approach to improve its performance. *Journal of Affective Disorders, 46*(1), 1-13.

- Geisser, M.E., Roth, R.S., & Robinson, M.E. (1997). "Assessing depression among persons with chronic pain using the Center for Epidemiological Studies-Depression Scale and the Beck Depression Inventory: a comparative analysis." *Clinical Journal of Pain*, 13(2), 163-70.
- Gilbody, S., Sheldon, T., & House, A. (2008). Screening and case-finding instruments for depression: a meta-analysis. *Canadian Medical Association Journal*, 178(8), 997-1003.
- Golub, E.T., Latka, M., Hagan, H., Havens, J.R., Hudson, S.M., Kapadia, F., Campbell, J.V., et al. (2004). Screening for depressive symptoms among HCV-infected injection drug users: examination of the utility of the CES-D and the Beck Depression Inventory. *Journal of Urban Health Bulletin* 81(2), doi:10.1093/jurban/jth114
- Goncalves, G., & Fagulha, T. (2004). The portuguese version of the center for epidemiologic depression scale (CES-D). *European Journal of Psychological Assessment*, 20(4), 339-348.
- 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*, 2, 485-508.
- Grossi, E., Mosconi, P., Groth, N., Niero, M., & Apolone, G. (2002). *Questionario Psychological Geeral Well-Being Index. Versione Italiana*. Milano: Istituto di Ricerche Farmacologiche "Mario Negri".
- Haringsma, R., Engels, G.I., Beekman, A.T.F., & Spinhoven, P. (2004). The criterion validity of the Center for Epidemiological Studies Depression Scale (CES-D) in a sample of self-referred elders with depressive symptomatology. *International Journal of Geriatric Psychiatry*, 19(6), 558-563.
- Iacoviello, M.B., Alloy, L.B., Abramson, L.Y., & Choi, J.Y. (2010). The early course of depression. A longitudinal investigation of prodromal symptoms and

- their relation to the symptomatic course of depressive episodes. *Journal of Abnormal Psychology*, 119(3), 459–467.
- Julian, L.J., Gregorich, S.E., Tonner, C., Yazdany, J., Trupin, L., Criswell, L.A., Yelin, E., et al. (2011). Using the Center for Epidemiologic Studies Depression Scale to screen for depression in systemic lupus erythematosus. *Arthritis Care Research*, 63(6), 884-890.
- Katon, W., & Schulberg, H. (1992). Epidemiology of depression in primary care. *General Hospital Psychiatry*, 14(4), 237-247.
- Kerr, L.K., & Kerr, L.D. (2001). Screening tools for depression in primary care. *West Journal of Medicine*, 175(5), 349-352.
- Keyes, C.L.M., Shmotkin, D., & Ryff, C.D. (2002). Optimizing well-being: the empirical encounter of two traditions. *Journal of Personality and Social Psychology*, 82(6), 1007–1022.
- Kim, G., Decoster, J., Huang, C.H., & Chiriboga, D.A. (2011). Race/ethnicity and the factor structure of the Center for Epidemiologic Studies Depression Scale: A meta-analysis. *Cultural Diversity Ethnic Minority Psychology*, 17(4), 381-96.
- Lewinsohn, P.M., Seeley, J.R., Roberts, R.E., & Allen, N.B. (1997). Center for Epidemiologic Studies Depression Scale (CES-D) as a screening instrument for depression among community-residing older adults. *Psychology and Aging*, 12(2), 277-287.
- Li, Z., & Hicks, M.H.R. (2010). The CES-D in Chinese American women: construct validity, diagnostic validity for major depression, and cultural response bias. *Psychiatry Research*, 175(3), 227-232.
- Logsdon, M.C., McBride, A.B., & Dean, A. (1994). Social support and postpartum depression. *Research in Nursing and Health* 17(6), 449-457.

- Lyubomirsky, S., Sheldon, K.M., & Schkade, D. (2005). *Pursuing happiness: the architecture of sustainable change*. UC Riverside.
- Madianos, M.G., Gournas, G., & Stefanis, C.N. (1992). Depressive symptoms and depression among elderly people in Athens. *Acta Psychiatrica Scandinavica*, 86, 320-326.
- Manhert, F., Reicher, H., Zalaudek, K., & Zapotoczky, H. (1997). Prodromal and residual symptoms in recurrent depression. *European Neuropsychopharmacology*, 7(2), 159–160.
- Marsh, H.W. (1986). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22(1), 37-49.
- Martens, M.P., Parker, J.C., Smarr, K.L., Hewett, J.E., Ge, B., Slaughter, J.R., & Walker, S.E. (2006). Development of a Shortened Center for Epidemiological Studies Depression Scale for Assessment of Depression in Rheumatoid Arthritis. *Rehabilitation Psychology*, 51(2), 135-139.
- Massé, R. (2000). Qualitative and quantitative analyses of psychological distress: methodological complementarity and ontological incommensurability. *Qualitative Health Research*, 10, 411-423.
- Mehrens, W., & Lehmann, I. (1983). *Measurement and evaluation in education and psychology* (3rd ed.). New York: Holt, Rinehart & Winston.
- Morin, A.J.S., Moullec, G., Maïano, C., Layet, L., Just, J.L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and nonclinical adults. *Revue Épidémiologie et de Santé Publique*, 59(5), 327-40.
- Myers, J.K., & Weissman, M.M. (1980). Use of a self-report symptom scale to detect depression in a community sample. *The American Journal of Psychiatry*, 137(9), 1081-1084.

- Pandya, R.B., Metz, L., & Patten, S.B. (2005). Predictive value of the CES-D in detecting depression among candidates for disease-modifying multiple sclerosis treatment. *Psychosomatics*, *46*(2), 131-134.
- Parikh, R.M., Eden, D.T., Price, T.R. & Robinson, R.G. (1988). The sensitivity and specificity of the Center for Epidemiologic Studies Depression Scale in screening for post-stroke depression. *The International Journal of Psychiatry in Medicine*, *18*(2), 169-181.
- Pilotte, W.J., & Gable, R.K. (1990). The impact of positive and negative itemstems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*, 603-610.
- Preamble to the Constitution of the World Health Organization as adopted by the International Health Conference, New York, 19-22 June, 1946; signed on 22 July 1946 by the representatives of 61 States (Official Records of the World Health Organization, no. 2, p. 100) and entered into force on 7 April 1948.
- Radloff, L.S. (1977). The CES-D Scale. *Applied Psychological Measurement*, *1*(3), 385-401.
- Radloff, L.S., & Locke, B.Z. (1986). The Community Mental Health Assessment Survey and the CES-D Scale. In M.M. Weissman, J.K. Myers, & C.E. Ross (Eds). *Community Surveys of Psychiatric Disorders*, 177-189.
- Ridner, S.H. (2004). Psychological distress: concept analysis. *Journal of Advanced Nursing*, *45*(5), 536-545.
- Roberts, R.E., Lewinsohn, P.M., & Seeley, J.R. (1991). Screening for adolescent depression: a comparison of depression scales. *Journal of the American Academy of Child & Adolescent Psychiatry*, *30*(1), 58-66.
- Roberts, R.E., & Vernon, S.W. (1983). The Center for Epidemiologic Studies Depression Scale: its use in a community sample. *The American Journal of Psychiatry*, *140*(1), 41-46.

- Ros, L., Latorre, J.M., Aguilar, M.J., Serrano, J.P., Navarro, B., & Ricarte, J.J. (2011). Factor structure and psychometric properties of the center for epidemiologic studies depression scale (CES-D) in older populations with and without cognitive Impairment. *International Journal Of aging human development*, 72(2), 83-110.
- Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self-esteem and specific self-esteem: different concepts, different outcomes. *American Sociological Review*, 60(1), 141-156.
- Ryan, R.M., & Deci, E.L. (2001). On happiness and human potentials: A Review of Research on Hedonic and Eudaimonic Well-Being. *Annual Review of Psychology*, 52, 141–66.
- Ryff, C.D. (1982). Successful aging: a developmental approach. *The Gerontologist*, 22, 209-214.
- Ryff, C.D. (1985). Adult personality development and the motivation for personal growth. In D. Kleiber & M. Maher (Eds.), *Advances in motivation and achievement: Motivation and adulthood* (Vol. 4, pp. 55–92). Greenwich, CT: JAI Press.
- 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, 1069–81.
- Ryff, C.D., & Keyes, C.L.M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology*, 69, 719–727.
- Ryff, C.D., & Singer, B.H. (1996). Psychological well-being: Meaning, measurement, and implications for psychotherapy research. *Psychotherapy and Psychosomatics*, 65, 14–23.
- Ryff, C.D., & Singer, B. (1998). The contours of positive human health. *Psychological Inquiry*, 9, 1–28.

- Schriesheim, C.A., Eisenbach, R.J., & Hill, K.D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, 51, 67-78.
- Schriesheim, C.A., & Hill, K.D. (1981). Controlling acquiescence response bias by itemreversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, 41, 1101-1114.
- Schulberg, H.C., Saul, M., McClelland, M., Ganguli, M., Christy, W., & Frank, R. (1985). Assessing depression in primary medical and psychiatric practices. *Archives of General Psychiatry*, 42(12), 1164-1170.
- Sharp, K.L., & Lipsky, M.S. (2002). Screening for depression across the lifespan: a review of measures for use in primary care settings. *American Family Physician*, 66(6), 1001-1009.
- Shinar, D., Gross, C.R., Price, T.R., Banko, M., Bolduc, P.L., & Robinson, R.G. (1986). Screening for depression in stroke patients: the reliability and validity of the Center for Epidemiologic Studies Depression Scale. *Stroke*, 17(2), 241-245.
- Soler, J., Pérez-Sola, V., Puigdemont, D., Pérez-Blanco, J., Figueres, M., & Alvarez, E. (1997). Validation study of the Center for Epidemiological Studies-Depression of a Spanish population of patients with affective disorders. *Actas Lusoespanolas de Neurologia Psiquiatria y Ciencias Afines*, 25(4), 243-249.
- Spijker, J., Van Der Wurff, F.B., Poort, E.C., Smits, C.H.M., Verhoeff, A.P., & Beekman, A.T.F. (2004). Depression in first generation labour migrants in Western Europe: the utility of the Center for Epidemiologic Studies Depression Scale (CES-D). *International Journal of Geriatric Psychiatry*, 19(6), 538-544.
- Springer, K.W., & Hauser, R.M. (2006). An assessment of the construct validity of Ryff's scale of psychological well-being: Method, mode, and measurement effects. *Social Science Research*, 35(4), 1080-1102.

- Stahl, D., Sum, C.F., Lum, S.S., Liow, P.H., Chan, Y.H., Verma, S., Chua, H.C., et al. (2008). Screening for depressive symptoms: validation of the center for epidemiologic studies depression scale (CES-D) in a multiethnic group of patients with diabetes in Singapore. *Diabetes Care*, *31*(6), 1118-1119.
- Suh, E.M. (1999). Culture, identity, consistency, and subjective well-being. Unpublished doctoral dissertation. Urbana-Champaign: University of Illinois.
- Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). Interanal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using dat from the Scottish Health Education Population Survey. *Health and Quality of Life Outcome*, *7*, 15, doi:10.1186/1477-7525-7-15
- Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., Parkinson, J., Secker, J., & Stewart-Brown, S. (2007). The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. *Health and Quality of Life Outcomes*, *5*(1), 63, doi:10.1186/1477-7525-5-63
- Thombs, B.D., Hudson, M., Schieir, O., Taillefer, S.S., & Baron, M. (2008). Reliability and validity of the center for epidemiologic studies depression scale in patients with systemic sclerosis. *Arthritis & Rheumatism*, *59*(3), 438-443.
- Vázquez, F.L., Blanco, V., & López, M. (2007). An adaptation of the Center for Epidemiologic Studies Depression Scale for use in non-psychiatric Spanish populations. *Psychiatry Research*, *149*(1-3), 247-252.
- Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M., et al. (1983). Development and validation of a geriatric depression screening scale: a preliminary report. *Journal of Psychiatric Research*, *17*, 37-49.
- Zich, J.M., Attkisson, C.C., & Greenfield, T.K. (1990). Screening for depression in primary care clinics: the CES-D and the BDI. *The International Journal of Psychiatry in Medicine*, *20*(3), 259-77.

- 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.
- Weissman, M.M., Sholomskas, D., Pottenger, M., Prusoff, B.A., & Locke, B.Z. (1977). Assessing depressive symptoms in five psychiatric populations: a validation study. *American Journal of Epidemiology*, 106(3), 203-214.
- World Health Organisation (2004). *Promoting Mental health; Concept emerging evidence and practice. Summary report*. Geneva: World Health Organisation.
- World Health Organisation (2001). *Strengthening mental health promotion. Summary report*. Geneva: World Health Organisation.
- Wright, B., & Masters, M. (1982). *Rating scale analysis*. Chicago: Mesa Press.