Rasch Analysis and Confirmatory Factor Analysis for the Flourishing Model with Masters Athletes: Study from France

*Dr/ Marei Salama-Younes

Introduction

Flourishing generally refers to the good life—feeling good and functioning effectively (Huppert & So, 2013). Through there are no unique or universal definitions of well-being and other positive psychological functioning (Coleman, 2009). Numerous models of flourishing have been proposed. For example, (Seligman, 2012) introduced the PERMA model. Huppert and So (2013) suggested 10 components of flourishing that are the opposite of depression and anxiety. (Diener et al., 2010) developed a brief 8-item measure of flourishing, designed to assess psychosocial success across several areas of life and to complement measures that focus on affect and life satisfaction. (Keyes, 2010) has developed a bipolar model for flourishing.

A growing body of research challenges the assumption by which mental illness (psychological distress) and mental health (well-being) are two separate, albeit related, continua (Keyes, 2005). Population-based studies investigating the prevalence and characteristics of mental health, in addition to mental illness, are therefore vital for providing evidence to support effective population intervention programs (Lamers, Westerhof, Bohlmeijer, ten Klooster, & Keyes, 2011). As proposed by Keyes and his colleagues (Keyes, 2007; Keyes & Waterman, 2003), follows the multidimensional model of mental health (flourishing model) includes both emotional and functional well-being as important domains of mental health.

These two domains reflect concepts identified by

* Department of Psychology, Sociology and Evaluation in Sport Field, Helwan University, Cairo, Egypt, & CRPCC-LAUREPS, Université de Rennes II and CREAD-IUFM de Bretagne, France.

many writers like Ryan and Deci (2001). More specifically, emotional well-being reflects the Greek concept of hedonic well-being (i.e., happiness or pleasure in life). It is comprised of avowed quality of life (viz., happiness and satisfaction with life) and positive affect, which parallels Diener, Oishi, and Lucas (2003) construction of subjective well-being. Functional well-being reflects the Greek concept of eudaimonic well-being, that is, feelings experienced when engaging in behaviors aimed at reaching one's potential (Diener et al., 1985). It consists of both the social and psychological well-being components.

Following this model, two competing models were tested. The single factor model hypothesizes that the measures of mental health and mental illness reflect a single latent factor, support for which would indicate that the absence of mental illness implies the presence of mental health. The two factor model hypothesizes that the measures of mental illness represent the latent factor of mental health that is distinct from, but correlated with, the latent factor of mental illness that is represented by the measures of mental illness (Keyes, 2005). The data strongly supported the two factor model, which was a nearly perfect fitting model to the (MIDUS) - a study on midlife development in the United States on a survey of 3,032 American adults- data (Keyes, 2005). As predicted, there is a modest association between mental health and mental illness; level of mental health tends to increase as level of mental illness decreases. The modest correlation suggests, however, that the latent constructs of mental health and mental illness are distinctive.

Support for the two factor model provides the strongest scientific evidence to date in support of the complete health approach to mental health. The evidence indicates that the absence of mental illness does not imply the presence of mental health, and the absence of mental health does not imply the presence of mental illness. Thus, neither the pathogenic nor salutogenic approaches alone accurately describe the mental health of a population. Rather, mental health is a complete state that is best studied through the combined assessments of mental health with mental illness. Complete mental health is a state in which individuals are free of mental illness and they are flourishing. Of course, flourishing may sometimes occur with an episode of mental illness, and moderate mental health and languishing can occur both with and without a mental illness.

In literature, (Keyes, 2007) has constructed the Mental Health Continuum-Short Form (MHC-SF). It consists of three well-being sub-scales served as indicators of positive mental health (Keyes, 2007): the summed scale of emotional well-being (i.e., single item of happiness, single item of life satisfaction and single item of positive affect), the summed scale of social well-being (i.e., single item of the fives sub-scales added together) and the summed scale of psychological well-being (i.e., single item of the six sub-scales added together), and Thus, Keyes' model of mental health consists of three broad domains: emotional, social and psychological well-being, which can be further subdivided into 14 specific domains. Each measure of subjective well-being is considered a symptom or characteristic of mental health insofar as it represents an outward sign of an otherwise unobservable state of mental health.

More precisely, mental health can be seen as a continuum, where an individual's mental health may have many different possible values (Keyes, 2002). The theory that the measures of mental health and mental illness belong to latent continua was tested using data from a representative sample of American adults. Three scales served as indicators of mental health composed of emotional, social and psychological well-being. Four summary measures served as indicators of mental illness, based on the number of symptoms of four mental disorders: generalized anxiety, panic disorder, major depressive episode, and alcohol dependence (Keyes et al., 2008).

Two competing theories were tested. The single factor model hypothesizes that the measures of mental health and mental illness reflect a single latent factor, support for which would indicate that the absence of mental illness implies the presence of mental health. The two-factor model hypothesizes that the measures of mental illness represent the latent factor of mental health that is distinct from, but correlated with the latent factor of mental illness that is represented by the measures of mental illness. The data strongly supported the two-factor model, which was a nearly perfect fitting model to the American findings (Keyes, 2005). As predicted, there is a modest association between mental health and mental illness; level of mental health tends to increase as level of mental illness decreases. However, the modest correlation suggests that the latent constructs of mental health and mental illness are distinctive. The evidence indicates that the absence of mental illness does not imply the presence of mental health, and the absence of mental health does not imply the presence of mental illness. Rather, mental health is a complete state that is best studied through the combined assessments of mental health with mental illness. Complete mental health is a state, in which individuals are free of mental illness and they are flourishing. Of course, flourishing may sometimes occur with an episode of mental illness, and moderate mental health and languishing can occur both with and without a mental illness.

In previous studies, for French seniors, the positive mental health, subjective vitality and life satisfaction were positively correlated but they were negatively correlated with the psychological distress (M Salama-Younes & Ismaïl, 2011). In addition, recently, there has been an increase in interest in issues related to the enhancement of the performance of the master's athlete. In France, there are no studies that evaluate the flourishing model for seniors' athletes (Mareï Salama-Younes, Ismaïl, Montazeri, & Roncin, 2011). There was only one study explored the complete mental health model with physically active old adult (Mareï Salama-Younes). However, there are not any study which explored the flourishing model as a bipolar construct for masters' athletes. In the current research, there is the main goal of the the first study. More precisely, two studies were conducted. In the first, the purpose was to test the flourishing model using the Confirmatory Factor Analysis

(CFA) and the Rasch analysis. In addition, the correlation between the two construct must then be performed. In the second study, if scales have been reduced, the confirmatory analysis should again be performed to confirm the new structure of the scale which be reduced.

Study 1

Participants and procedures

The participants of this study were French masters Athletes. (n =601) from Rennes city. It consisted of 317 women and 284 men. They aged from 35 to 48 years (M= 41.98±6.67 SD). The study has been administered in period from 19/03/2012 to 21/12/2012. Sample practiced only one of the following sport activities Cycling (289), Running (166) or Swimming (146). Currently they were practicing their activity 3 or 4 times per week. Researchers informed the participants about the objective of the study and that their participation was voluntary and they could withdraw at any time. Both oral and written instructions were given regarding items understanding (i.e., that there were no right or wrong answers to the questions and they should freely state what they think), and they were reassured about the confidentiality of their responses.

Measures

Short Flourishing Scale (MHC-SF). (Keyes, 2007) The MHC–SF consists of 14 items. An adapted version of MHC-SF is used where it use a 6-point Likert-type scale (from 1 to 6). It measures the degree of (a) emotional well–being (EWB) (items 1-3) as defined in terms of positive affect/satisfaction with life; (b) social well–being (SWB) (items 4-8) as described in Keyes' (1998) model of social well–being (one item on each of the facets of social acceptance, social actualization, social contribution, social coherence, and social integration); and (c) psychological well–being (PWB) (items 9-14) as described in Ryff's and Keyes's (1995) model (including one item on each of the dimensions of autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self–acceptance). Short version has been used in different cultures (Gilmour, 2014; Karaś, Cieciuch, & Keyes, 2014; Keyes et al., 2008; Lamers et al., 2011).

General Health Questionnaire (GHQ-12). (Goldberg & Williams, 1988) The GHQ-12 assesses the severity of a mental problem over the past few weeks. We used an updated version of GHQ-12 of 4-point Likert-type scale (from 0 to 3). The score was used to generate a total score ranging from 0 to 36. This questionnaire was validated in many countries and languages (López & Dresch, 2008; Montazeri et al., 2003; Politi, Piccinelli, & Wilkinson, 1994; Mareï Salama-Younes, Montazeri, Ismaïl, & Roncin, 2009). Moreover, it has rarely been used with athletes (Mareï Salama-Younes, et al., 2009)

Statistical Analysis

Confirmatory factor analysis (CFA) is a procedure for learning the extent to which **k** observed variables might measure **m** abstract variables, where in m is less than k. In CFA, we indirectly measure non-observable behavior by taking measures on multiple observed behaviors. Conceptually, in using CFA we can assume either nominalist or realist constructs, yet most applications of CFA in the social sciences assume realist constructs (Harrington, 2008).

Rasch analysis. Use of Rasch model methodology involves a rigorous and extensive analysis of the data and provides additional psychometric information that cannot be obtained through the Classical Test Theory (CTT) approach. In Rasch model analysis, the goal is to evaluate the degree to which each of the scale items and, in turn, item categories fit within a mathematical model. This, in short, suggests that responses to individual items should be predictable from the individual's overall scale score, and vice versa. In so doing, evidence for validity of the scale is provided through fit indicators and response thresholds. These results could be viewed as complementary to classical approaches – i.e., CTT, both of which contribute to evidence for validity of the scale.

Otherwise, Rasch analysis is considered one of the most suitable methods to achieve this aim as it provides advanced techniques for examining psychometrical properties of scales. Accordingly this may reduce the number of items scale for these differences constructs or require some further actions (Beaton, Wright, & Katz, 2005). Rasch model, as one of IRT models, comprises a collection of modeling techniques for the analysis of item level data obtained to measure inter-individual variation. This collection of techniques generates rich item level information and offers many advantages over classical test theory (CTT) (Meijer & Nering, 1999; Prieto, Alonso, & Lamarca, 2003). Rasch model is used to evaluate the psychometric properties of an existing scale and its items to evaluate its performance. When used appropriately, Rasch modeling can produce precise, valid, and relatively instruments (Edelen & Reeve, 2007; Prieto et al., 2003).

Rasch analysis and sample size requirement. Although there is no definitive answers regarding sample size requirements for IRT models, there are some general statements and guidelines that can be outlined. Sample size needs increase with the complexity of IRT model. Sample sizes as small as 100 are often adequate for estimating stable Rasch-model parameters (Linacre, 1994). In this study, a large sample composed of 601 is used to apply Rasch model analysis with stable model parameters.

The final samples for each scale excluded persons scoring at the floor and ceiling levels because their item responses do not vary and their standard errors are infinite (Linacre, 2012). Person-and-item-fit statistics were examined for departures from model expectations. The following analyses were performed: (a) constructing the person–item map; (b) testing of fit between the data and the model; (c) estimating the person and item reliability coefficient; (d) testing the ordering of the categories; (e) analyzing the dimensionality; (f) converting the logits scale to more meaningful scale.

Person–item map. A map was constructed of the hierarchy of the person and item measures to examine item and person performances. At the bottom of the map, the lower estimates of the person and item can be found, with increasing estimates represented higher up the map. On the left side, the persons' performances are represented and on the right side the items. For a well-targeted measure, the mean location for the person should be around zero logits and the items' measures should be near the corresponding persons' measures to make sure that persons were assessed with suitable items measures (John M Linacre, 2012)

Test of fit to the model. Determining how well the empirical data fit the Rasch model is performed through the items fit statistics. These fit statistics are the infit mean square (infit) and the outfit mean square (outfit) for estimating person's measure (i.e., ability) and item's measure (i.e., difficulty- also called the b parameter). The infit and the outfit represent the information-weighted mean square residual difference between observed and expected responses. The infit statistics are sensitive to unexpected responses near the person's ability level and the outfit statistic is more sensitive to outliers. The optimal expected infit or outfit mean square values are 1.0 (John M Linacre, 2012). One of the important aspects to examine is the standard error of measurement (SEM) for items' and persons' measures. There is no exact criteria for judging the SEM but the less the standard error is, the more is the precise is the measures (Baker, 2001). The model overall fit is also achieved when items can be modeled with ICC of the IRT model used, i.e., Rasch model, and overall is modelled by test characteristic curve (TCC) of the same model used (John M Linacre, 2012; Orlando, Sherbourne, & Thissen, 2000). Usually, Rasch analysis starts by examining persons' fitness and then items' fitness and ends by examining the TCC.

Reliability statistics. In the Rasch model, reliability is estimated for both persons and for items. Person reliability in WINSTEPS is equivalent to the test reliability (Cronbach's alpha) in the CTT. The person reliability reports how reproducible the person's ability order is in this sample of persons for this set of items. The item reliability reports how reproducible the item's difficulty order is for this set of items for this sample of persons (Linacre, 2012).

Category function. Category functioning is examined by analyzing category frequencies, mean measures, thresholds, and category fit statistics. The category frequencies indicate how many persons chose a particular response category. The recommended minimal number of responses per category is ten for stable rating scale–structure threshold parameter (Linacre, 2012). The mean measures and the thresholds should increase when moving from lower to higher categories. When there are ordered categories, the category probability curves show that each category is the most probable category at some point on the latent variable the scale measures (John M Linacre, 2012).

Dimensionality investigation. One of Rasch model important assumptions is unidimensional. When the data fits the Rasch model, the Rasch dimension is the only dimension in the data. However, using usual factor analysis may be not the right test for examining Rasch model unidimensionality (John M Linacre, 2012) .A Rasch-residual-based Principal Components Analysis (PCAR) is as a usual factor analysis, but the components show contrasts between opposing factors, not loadings on one factor. Although criteria have not been established for when a deviation becomes a dimension, PCA is indicative, but not definitive, about secondary dimensions. Unidimensionality is accepted if the Rasch dimension explains > 50% of the variance of measures in the data and the largest secondary dimension, "the first contrast in the residuals" explains < 3.0 (Linacre, 2012).

Converting the logits scale to more meaningful reporting scale. As persons' ability scale is in logits which ranges from -4 to 4 in most cases, it is always preferred to convert the scores to more meaningful reporting scale to avoid negative values and decimal points using linear transformation (Kazem & El-Shiekh, 2000). Persons' ability were estimated in logits, and then transformed and rounded into a reporting scale with mean 50 and standard deviation 10.

The study data was managed and analyzed by using SPSS (IBM, 2012), LISREL vs 8.7 and Rasch analysis was performed by using WINSTEPS (John Michael Linacre, 2006).

Results

Firstly, the internal consistency was assessed by calculating the Cronbach's coefficient. The values of 0.70 or greater were considered satisfactory. After being tested the factor structure by EFA, we performed CFA to assess the structure of two scales. The intention was to indicate if the model fits well the data. There are varying suggestions in the literature about the number, type, and cut-off values for goodness-of-fit required to be reported for CFA (Byrne, 2009). A popular recommendation is to present three of four indices from different areas. Accordingly, we report several goodness-of-fit indicators including GFI (Goodness-of-Fit Index), NFI (Normed Fit Index), RMR (Root Mean Square Residual), RMSEA (Root Mean Square Error of Approximation), and $\chi 2$/df (see table 1 and figure 1). The recommended cut-off values for acceptable values are ≥ 0.90 for GFI and NFI. The RMR and RMSEA test the fit of the model to the covariance matrix. As a guideline, values below 0.05 indicate a close fit and values below 0.11 are an acceptable fit. The value of 2 alone may be used as an index, but 2 divided by the degrees of freedom (2/df) reduces its sensitivity to sample size (cut-off values < 2-5). For masters athletes, the GFIs for MHC-SF and GHQ-12 were acceptable in terms of $\chi 2$/df ratio, GFI, RMR and RMSEA.

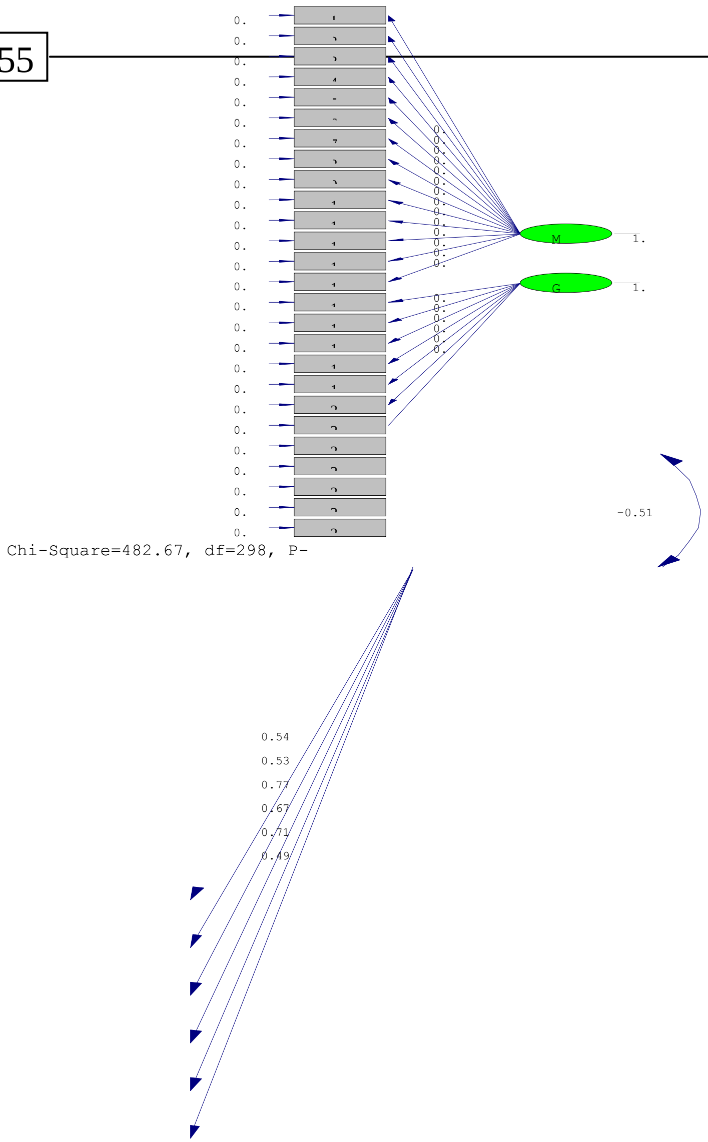
Table (1)

Goodness-of-fit of the confirmatory factor analysis models n = 601

Scales	$\chi 2$	df	GFI	NFI	RMR	RMSEA	$\chi 2$ / df	(α)
MHC-SF (14 item)	256.78	74	0.92	0.90	0.07	0.071	4.58**	0.84
MHC-SF (11 item)	107.42	44	0.97	0.94	0.01	0.006	1.96**	0.80
GHQ-12 (12 item)	117.52	51	0.96	0.92	0.04	0.034	1.96**	0.84

Note:

GFI = Goodness of Fit Index, NFI = Normed Fit Index, RMR = Root Mean Square Residual, RMSEA = Root Mean Square Error of Approximation, * $p < .01$; ** $p < .001$.


M
G
1.
1.
-0.51
Chi-Square=482.67, df=298, P-
0.54
0.53
0.77
0.67
0.71
0.49

The MHC-SF and GHQ-12 scales data were then transferred into the Rasch Rating Scale Model (RSM) (John M Linacre, 2012) after reversing negative items. Following the Rasch model analysis procedures, each scale results will be reported separately as follows:

Mental Health Continuum-Short Form (MHC-SF)

The relationship between persons and items as shown in, figure 2 shows that the items match the persons' ability range. Concerning the persons' fit, 36 persons were deleted during the scaling analysis as they were poorly fitting. For the items, three items (items: 5, 6 and 13) were found misfit and were deleted (Table 2). Finally, the new scale test characteristic curve followed the Rasch model (Figure 3a).The new scale mean was .00±1.91 with .08±.01 standard error. The persons and items reliability were .86 and .99 respectively. MHC-SF scale category proved to be fitting to the model as it had acceptable frequencies (126-854), monotonically increasing threshold (-1.72 to 1.91) and none of the categories showed a misfit. Rasch model explained 77.21% of the variance of measures and the first contrast in the residuals was only 2. The rounded transformed persons' ability range on the modified MHC-SF was (24-88) and mean 67.26±08.29 with SEM 2.99±.77.

Table (2)

Means and standard deviations of the study scales in logits

Item No.	MHC-SF		GHQ-12	
	Measure	S.E	Measure	S.E
1	-1.34	.08	-1.00	.14
2	-1.43	.08	1.17	.11
3	-.79	.07	-1.17	.16
4	1.11	.08	-1.18	.13
5	×	×	.46	.11
6	×	×	.89	.05
7	1.91	.07	-1.62	.08

Follow Table (2)
Means and standard deviations of the study scales in logits

Item No.	MHC-SF		GHQ-12	
	Measure	**S.E**	**Measure**	**S.E**
8	1.45	.08	-0.94	.21
9	-.69	.08	1.87	.21
10	-1.03	.07	1.06	.03
11	-1.72	.07	1.38	.13
12	.94	.07	-1.54	.11
13	×	×	-	-
14	-.19	.08	-	-
Mean	**.00**	**.07**	**.00**	**.57**
S.D.	**1.27**	**.02**	**1.19**	**.14**
Person Reliability*	.86	.83		
Item Reliability	.99	.99		

Note. * is equivalent to the traditional "test" reliability; × item was deleted during the analysis; - not applicable.

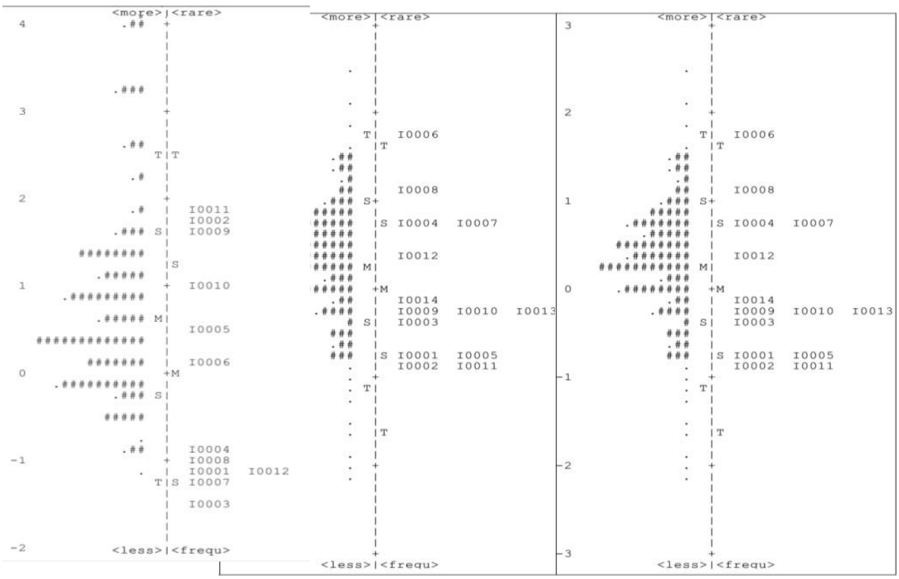


Figure 2 **Person-item maps for the study instruments**

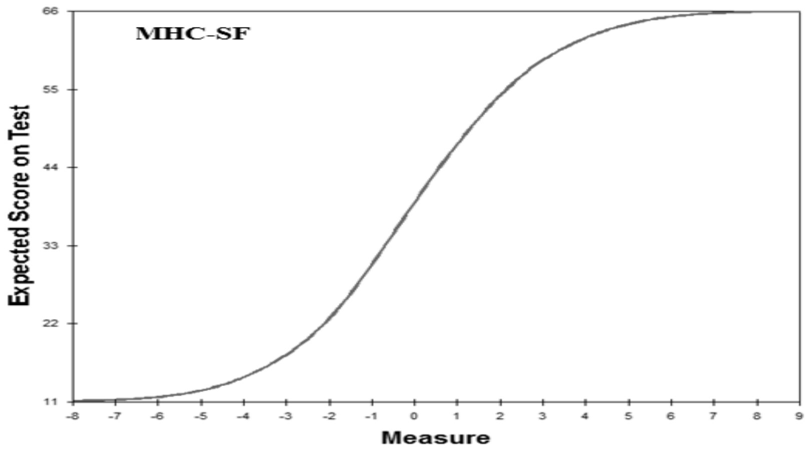


Figure 3a**, Test characteristics curves (TCC) for study instruments (MHC-SF)**

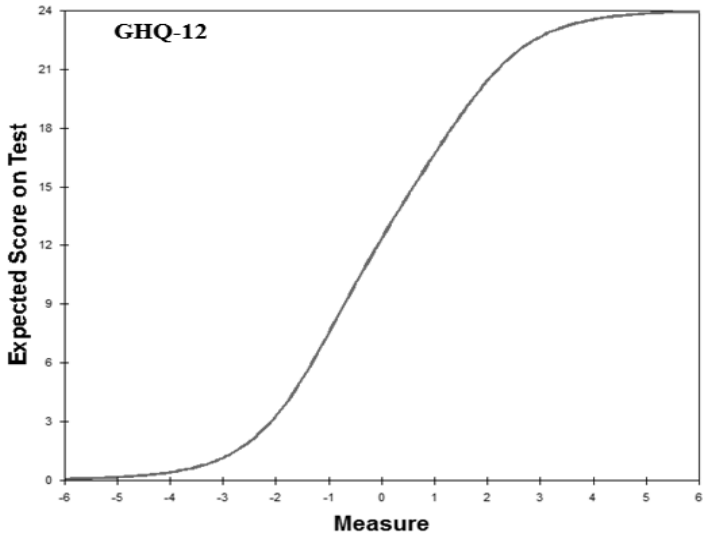


Figure 3 b, **Test characteristics curves (TCC) for study instruments (GHQ-12)**
General Health Questionnaire (GHQ-12)

As none of the current sample choose the category (3) of the response on GHQ-12 items, except for items 3 and 9 which had only twelve persons chose this response. Therefore, items were re-scored to be from 0,1,2,3 into 0, 1, and 2 only before performing any Rasch analysis. The relationship between persons and items, as shown in figures 2and 3b, showed that the items matched the persons' ability range although many of items were gathered on the low area of the scale. Concerning the persons' fit, 82 persons were deleted during the scaling analysis as they were poorly fitting. For the items, no items was found misfit. Finally, the new scale test characteristic curve followed the Rasch model. The new scale mean was .00±1.37 with .55±.11 SEM. The persons and items reliability were .70 and .99 respectively. GHQ-12 scale category proved to be fitting to the model as it had acceptable frequencies (674-1284), and monotonically increasing threshold (-1.62 to 1.87). Rasch model explained 63.13% of the variance of measures and the first contrast in the residuals was only 3.22. The rounded transformed persons' ability range on the modified GHQ-12 was (31-92) and mean 67.28±12.33 with SEM 4.32±0.92.

The relationship among the study variables

Conversion tables – i.e., from total raw scores to scaled scores - were estimated using WINSTEPS. After total raw scores of each instrument was converted to scale scores in Logits and then to reporting score, Pearson correlations were computed between the total scale scores for all sample (n=601); i.e., miss fitting persons were deleted only during scaling and were used during computing correlations. As shown in Table 2, there were negative significant correlations between GHQ-12 and MHC-SF (r= -.51; p< .001)

. **Study 2**

Participants and procedures

The participants of this study were French masters Athletes. (n =337) from Rennes city. It consisted of 177 women and 160 men. They aged from 35 to 45 years (M= 40.09±4.43 SD). The study has been administered in period from 11/04/2013 to 22/06/2013. Participants practiced Cycling (137), Running (106) or Swimming (94). Currently they were practicing also their activity 3 or 4 times per week. Researcher informed them about the objective of the study and that their participation was voluntary and they could withdraw at any time. Both oral and written instructions were given regarding items understanding (i.e., that there were no right or wrong answers to the questions and they should freely state what they think), and they were reassured about the confidentiality of their responses.

Measures

Short Flourishing Scale (MHC-SF). It consists of 11 items. It measures the degree of (a) emotional well–being (EWB) (items 1-3) as defined in terms of positive affect/satisfaction with life; (b) social well–being (SWB) (items 4, 7, 8) and (c) psychological well–being (PWB) (items 9, 10, 11, 12, 14)

Statistical Analysis

The second study data was managed and analyzed by using SPSS (IBM, 2012) and LISREL vs 8.7.We performed CFA to assess the reduced structure for the short flourishing scale.

Results

The intention was to indicate if the model fits well the data. We report several goodness-of-fit indicators including GFI (Goodness-of-Fit Index), NFI (Normed Fit Index), RMR (Root Mean Square Residual), RMSEA (Root Mean Square Error of Approximation), and $\chi 2$/df (see figure 4). The recommended cut-off values for acceptable values are ≥ 0.90 for GFI and NFI. The RMR and RMSEA test the fit of the model to the covariance matrix. As a guideline, values below 0.05 indicate a close fit and values below 0.11 are an acceptable fit. The value of 2 alone may be used as an index, but 2 divided by the degrees of freedom (2/df) reduces its sensitivity to sample size (cut-off values < 2-5). For masters' athletes, the model consists of only 11 items fit good the data. The Short Flourishing Scale was acceptable in terms of $\chi 2$/df ratio, GFI, RMR and RMSEA. In complemtary with the CFA, the item response theory must be used in development, validation and adaptation of cross-culture scales.

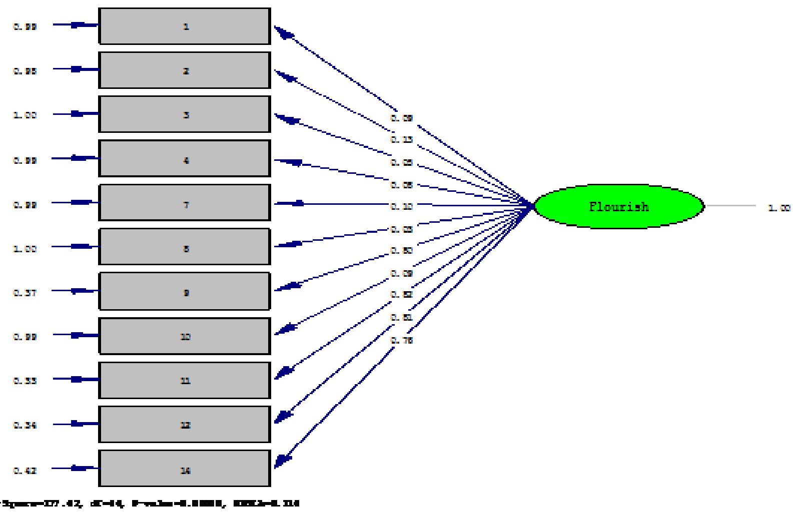


Figure 4, **CFA for the reduced Flourishing scale**

Conclusion and Discussion

This study basically aimed to test and examine the construct of the MHC-SF, and the GHQ-12 instruments using CFA and Rasch analysis methodology. Further to examining the scales construct, the relationships between the bipolar construct were also examined in terms of previous studies based on CTT. Although CTT used by other researchers is valid and reliable, but using the Rasch model in-depth analysis in exploring the well-being scales provided more efficient, reliable, and valid results as it provided additional psychometric information that cannot be obtain through the CTT approach.

Firstly, The data were firstly screened for nonnormality, and no problematic trend was detected. Univariate skewness ranged from -1.14 to 1.16, and univariate kurtosis ranged from -1.09 to 1.08, indicating that the responses were relatively normally distributed. In addition, relative multivariate kurtosis as reported by the output from LISREL 8.7 equalled 1.14. Whereas, there is no standard cut-off for this index, Hu and Bentler (1998) recommended that multivariate normality can be assumed if this value is less than 3. So, the distribution of variables being normal a matrix of product-moment correlations was generated to evaluate the models using the maximum robust-likelihood method (ML-robust). In sum, the structural validity of the flourishing model, as a bipolar construct, has been examined with masters athletes. The bipolarity construct has been tested using CFA performed with LISREL 8.7 software (Jöreskog, Sörbom, & Du Toit, 2001). The intention was to indicate if the model fits well the data. There are varying suggestions in the literature about the number, type, and cut-off values for goodness-of-fit required to be reported for CFA (Keyes et al., 2008). A popular recommendation is to present three of four indices from different areas. The indices reported a bipolarity of model as shown (Figure 1). A correlation between the two construct confirmed the negative relationships (-0.51, $p > 0.01$).

In the second step, the versions of study instruments were well examined and revised through Rasch analysis after deleting miss-fitting persons and items. The scales proved to be good as persons' and items' reliability coefficients were examined and found acceptable. Items categories were also examined and proved to be monotonically increasing without any miss-fitting categories. And finally dimensionality was examined and proved through PCAR; i.e., according to Rasch's concept of unidimensionality, each of the two instruments proved to be unidimensional and the secondary dimensions proved to be just deviations that were not up to create a secondary dimension. Thus, the four instruments performed adequately using the thorough and stringent Rasch analysis and the new versions can be used in future studies.

However, some points of weakness were found in the instruments after analysis, although not series but should be mentioned and discussed. Regarding the relation between persons and items (Figure 2), it was clear that there was a relative shift between the distribution of the persons' abilities and the items difficulties in GHQ-12. This shift can lead to less accurate estimation of items parameters that can affect applying this result to other samples because the accuracy of estimation is precisely estimated when the distribution of persons' ability (measures) is consistent to the corresponding distribution of items difficulties (measures) (Kazem, 1996). To avoid this point of weakness, we recommend to performed analysis on non-homogenous samples with variety of ability levels. Also, trying to find more fitting IRT models - 2 or 3 parameters models - can help in improving the items-and-person relation and the accuracy of calibration and SEM.

It was noticed that there are some gaps between items' measures (difficulties) along the GHQ; there is a gap between item 4 (b=-.73) and item 6 (b=.36) as the difference between the two items (2.02) was larger than the sum of their SEM (.29); in MHC-SF, there is a gap between item 10 (b=-.64) and item 12 (b=.18) as the difference between the two items (.46) was larger than the sum of their SEM (.30). If the difference between two adjacent items is larger than the sum of their SEM, then there is a real gap that requires to be filled by adding items in this level of scale (Massoud, 2010). We recommend further examination for these gaps and enrichment of the instruments with items that cover them.

Regarding the correlations between the study variables, as predicted, results showed a significant negative correlation between the flourishing scale MHC and psychological distress GHQ-12. This result is similar to previous studies (Mareï Salama-Younes & Salama-Younes; Winzer, Lindblad, Sorjonen, & Lindberg, 2014) .This result is also similar to previous studies confirming that the three subscales are negatively correlated with the satisfaction with life, the subjective vitality and the mental health (Joshanloo, Wissing, Khumalo, & Lamers, 2013).

To conclude, many in-depth psychometrical properties were examined regarding the two bipolar construct through use of Rasch analysis methodology that may lead to further studies. In addition, the relationships between the scales were found consistent with the results of literature studies. Although further work is required to improve the weakness found in the scales, our findings indicate the importance of using Rasch analysis in developing such types of instruments for the purpose of enhancement, refinement and precise results.

References

Beaton, D. E., Wright, J. G., & Katz, J. N. (2005). Development of the QuickDASH: comparison of three item-reduction approaches. The Journal of Bone & Joint Surgery, 87(5), 1038-1046.

Coleman, J. (2009). Well- being in schools: empirical measure, or politician's dream? Oxford Review of Education, 35(3), 281-292.

Diener, E., Oishi, S., & Lucas, R. E. (2003). Personality, culture, and subjective well-being: Emotional and cognitive evaluations of life. Annual review of psychology, 54(1), 403-425.

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D.-w., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. Social Indicators Research, 97(2), 143-156.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. Quality of Life Research, 16(1), 5-18.

Gilmour, H. (2014). Positive mental health and mental illness. Health reports, 25(9), 3.

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

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods, 3(4), 424.

Huppert, F. A., & So, T. T. (2013). Flourishing across Europe: Application of a new conceptual framework for defining well-being. Social Indicators Research, 110(3), 837-861.

Jöreskog, K. G., Sörbom, D., & Du Toit, S. (2001). LISREL 8: New statistical features: Scientific Software International.

Joshanloo, M., Wissing, M. P., Khumalo, I. P., & Lamers, S. M. (2013). Measurement invariance of the Mental Health Continuum-Short Form (MHC-SF) across three cultural groups. Personality and Individual Differences, 55(7), 755-759.

Karaś, D., Cieciuch, J., & Keyes, C. L. (2014). The Polish adaptation of the Mental Health Continuum-Short Form (MHC-SF). Personality and

Individual Differences, 69, 104-109.
Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. Journal of consulting and clinical psychology, 73(3), 539.
Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: a complementary strategy for improving national mental health. American psychologist, 62(2), 95.
Keyes, C. L. (2010). Flourishing: Wiley Online Library.
Keyes, C. L., & Waterman, M. B. (2003). Dimensions of well-being and mental health in adulthood.
Keyes, C. L., Wissing, M., Potgieter, J. P., Temane, M., Kruger, A., & van Rooy, S. (2008). Evaluation of the mental health continuum–short form (MHC–SF) in setswana- speaking South Africans. Clinical Psychology & Psychotherapy, 15(3), 181-192.
Lamers, S., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum- short form (MHC- SF). Journal of clinical psychology, 67(1), 99-110.
Linacre, J. M. (2006). WINSTEPS Rasch measurement computer program. Chicago: Winsteps. com.
Linacre, J. M. (2012). Facets computer program for many-facet Rasch measurement, version 3.70. 0. Beaverton: Winsteps. com.
López, M. P. S., & Dresch, V. (2008). The 12-Item General Health Questionnaire (GHQ-12): Reliability, external validity and factor structure in the Spanish population. Psicothema, 20(4), 839-843.
Meijer, R. R., & Nering, M. L. (1999). Computerized adaptive testing: Overview and introduction. Applied psychological measurement, 23(3), 187-194.
Montazeri, A., Harirchi, A. M., Shariati, M., Garmaroudi, G., Ebadi, M., & Fateh, A. (2003). The 12-item General Health Questionnaire (GHQ-12): translation and validation study of the Iranian version. Health and quality of life outcomes, 1(1), 66.
Orlando, M., Sherbourne, C. D., & Thissen, D. (2000). Summed-score linking using item response theory: Application to depression measurement. Psychological Assessment, 12(3), 354.
Politi, P., Piccinelli, M., & Wilkinson, G. (1994). Reliability, validity and factor structure of the12- item General Health Questionnaire among young males in Italy.

Acta Psychiatrica Scandinavica, 90(6), 432-437.
Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. Health and quality of life outcomes, 1(1), 27.
Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. Annual review of psychology, 52(1), 141-166.
Salama-Younes, M. Exploration on Keyes's Model of Mental Health for French Physically Active Adults, US China Education Journal, 3, 1, 933-943.
Salama-Younes, M., & Ismaïl, A. (2011). Validation of factor structure of mental health continuum (MHC–SF) for physically active older adults. World Journal of Sport Sciences, 4(1), 20-30.
Salama-Younes, M., Ismaïl, A., Montazeri, A., & Roncin, C. (2011). Factor structure of the French Ryff's Psychological Well-Being Scales for active older adults The human pursuit of well-being (pp. 203-212): Springer.
Salama-Younes, M., Montazeri, A., Ismaïl, A., & Roncin, C. (2009). Factor structure and internal consistency of the 12-item General Health Questionnaire (GHQ-12) and the Subjective Vitality Scale (VS), and the relationship between them: a study from France. Health and quality of life outcomes, 7(1), 22.
Salama-Younes, M., (2011). Positive Psychology, Character Strengths and Virtues (CSV): An Introduction, WCPP, July 23-26, Philadelphie, Pennsylvania, USA.
Seligman, M. E. (2012). Flourish: A visionary new understanding of happiness and well-being: Simon and Schuster.
Winzer, R., Lindblad, F., Sorjonen, K., & Lindberg, L. (2014). Positive versus negative mental health in emerging adulthood: a national cross-sectional survey. BMC public health, 14(1), 1238.