"Validation of the Mental Health Continuum-Short Form Revised in an English Speaking Student Population"

Alexander Dostall
Thesis Submitted for the Degree B. Sc.
Dr. Marion Sommers-Spijkerman
Prof. Dr. Gerben J. Westerhof
University of Twente
July 2019

Abstract

The Mental Health Continuum-Short Form Revised (MHC-SF-R) is the revised version of the Mental Health Continuum-Short Form (MHC-SF), a self-report questionnaire for assessing positive mental health. Since previous research revealed that various items in the MHC-SF are ambiguous and that the MHC-SF does not contain any item measuring relational well-being, the MHC-SF-R was developed for overcoming these shortcomings. In the MHC-SF-R the underlying structure of MHC-SF's emotional, psychological and social well-being dimension was taken over, the social well-being dimension was renamed into societal well-being and a fourth dimension investigating relational well-being was added. The primary aim of this study was to validate the MHC-SF-R by examining its factorial structure, internal consistency and convergent validity. The secondary aim was to investigate how the MHC-SF-R performs in comparison with the MHC-SF regarding psychometric properties. A cross-sectional questionnaire survey design was implemented and a convenient sample consisting of 107 English speaking students was utilized for analysis. For comparing MHC-SF-R's psychometric properties with MHC-SF's, the sample was split into two groups (MHC-SF-R/ MHC-SF). Against expectations, MHC-SF-R's model fit was not acceptable. Consistent with expectations, internal consistency for MHC-SF-R's total scale was high and convergent validity with related questionnaires was good. All of MHC-SF-R's subscales, except its societal well-being scale, displayed at least acceptable internal consistency indicating that the societal well-being scale needs further improvement. Due to the small sample, it was not possible to use the appropriate estimation method for conducting confirmatory factory analysis. Therefore, it is suggested to reexamine MHC-SF-R's model fit since in case future research reveals an acceptable model fit and internal consistency of MHC-SF-R's societal WB scale can be improved, the MHC-SF-R becomes a more comprehensive instrument for investigating positive mental health than the MHC-SF. However, for the time being the MHC-SF should be preferred over the MHC-SF-R.

Keywords: Mental Health Continuum-Short Form Revised, MHC-SF-R, mental well-being, psychometric properties, validation, factorial structure, validity, internal consistency, positive mental health, positive psychology

Table of Contents

1 Introduction

Traditionally, mental health models treat mental health and mental illness as two opposing poles situated on one continuum (Greenspoon & Saklofske, 2001). However, at present there is strong scientific evidence that psychopathology and positive mental health are not integral parts of one continuum but are distinct albeit related constructs contributing to mental health (Keyes, 2002, 2005b; Lamers, Westerhof, Bohlmeijer, Ten Klooster, & Keyes, 2011; Wang, Zhang, & Wang, 2011; Westerhof & Keyes, 2010). For this reason, it is important to have valid and reliable instruments not only for determining psychopathology but also for investigating positive mental health.

The Mental-Health-Continuum Long Form (MHC-LF) is a reliable and valid measure for assessing positive mental health (Keyes, 2005a). As the MHC-LF consists out of 40 items, it is a relative long questionnaire. Long questionnaires require longer periods of attention from respondents than short questionnaires. In clinical practice and research, usually more than one questionnaire is handed out for completion. Since periods of attention are limited (Kahneman, 1973), it is desirable that questionnaires are as short as possible under consideration of being valid and reliable. Therefore, MHC-LF's most informative items were evaluated and the Mental Health Continuum-Short Form (MHC-SF) consisting out of 14 items was developed (Keyes, 2002).

The MHC-SF (see Appendix A) consists of three items measuring emotional well-being (EWB), five items evaluating social well-being (SWB) and six items evaluating psychological well-being (PWB). EWB is based on the hedonic view of well-being which is devoted to focalize on happiness by defining well-being in terms of life satisfaction, the presence of positive emotions and the absence of negative emotions (Diener, Suh, Lucas, & Smith, 1999). In the MHC-SF, items assessing EWB measure happiness, interest in life and life satisfaction. SWB and PWB bear on the eudaimonic perspective of well-being. According to the eudaimonic perspective, well-being is defined in terms of the extent a person is fully positive functioning. Crucial aspects of it are self-realization and finding meaning in one's life (Keyes, 1998; Ryan & Deci, 2001). In the MHC-SF, assessment of SWB is inspired by Keyes's (1998) model of SWB which suggests that SWB consists of five dimensions, namely, social contribution, social integration, social actualization, social acceptance and social coherence. For determining SWB, the MHC-SF comprises one item for each of Keyes's (1998) dimensions. MHC-SF's PWB scale structure is derived from

Ryff's (1989) six dimensions of optimal functioning. Ryff (1989) subdivided PWB into self-acceptance, positive relations with others, autonomy, environmental mastery, purpose in life and personal growth. Here again, the MHC-SF contains one item for each dimension.

Despite a reduction from 40 to 14 items, the MHC-SF remains a reliable and valid instrument for measuring mental health in non-clinical (Echeverría et al., 2017; Keyes et al., 2008; Lamers et al., 2011; Perugini, Iglesia, Castro, & Keyes, 2017; Petrillo, Capone, Caso, & Keyes, 2015) and clinical samples (Franken, Lamers, Ten Klooster, Bohlmeijer, & Westerhof, 2018). Studies consistently showed that the MHC-SF fits best a three-factor structure composed of emotional, psychological, and social well-being. Additionally, those studies reported acceptable to excellent internal consistency for MHC-SF's total scale (Echeverría et al., 2017; Franken et al., 2018; Keyes et al., 2008; Lamers et al., 2011; Perugini et al., 2017; Petrillo et al., 2015). Convergent validity between MHC-SF's subscales and corresponding validation measures (life-satisfaction, individual functioning, involvement in society) is moderate (Keyes et al., 2008; Lamers et al., 2011; Perugini et al., 2017; Petrillo et al., 2015) and MHC-SF's discriminant validity with instruments measuring psychopathology is good (Lamers et al., 2011; Perugini et al., 2017). Moreover, Lamers et al. (2011) and Petrillo et al. (2015) found that the MHC-SF displays moderate test-retest reliability. Additionally, they suggest that the MHC-SF seems to be sensitive to change and simultaneously steady over time.

Although all subscales of the MHC-SF have sufficient reliability and validity, two subscales of the MHC-SF seem to perform less well, namely the SWB and PWB subscale. Across different populations, items loading on the SWB subscale display consistently lower factor loadings and smaller correlation values with theoretically relevant constructs compared to items loading on the PWB and EWB subscale. Moreover, items loading on the PWB subscale show consistently lower factor loadings compared to items loading on the EWB subscale (Lamers et al., 2011; Orpana, Vachon, Dykxhoorn, & Jayaraman, 2017; Petrillo et al., 2015; Vuletić, Erdeši, & Nikić, 2018). Those lower factor loadings might be explained by Köhle's (2010) finding that many SWB items and some PWB items seem to be ambiguous. By means of the method "thinking aloud", Köhle (2010) showed that citizens living in Enschede (the Netherlands) seem to have most difficulties interpreting SWB items followed by some PWB items. Support for the finding that SWB items seem to be most problematic comes from outpatients seeking treatment in the organization "Mindfit" located in Deventer (the Netherlands) who also reported most ambiguities while answering SWB items (Westerhof, 2019).

Therefore, a new version of the MHC-SF, the Mental Health Continuum-Short Form revised (MHC-SF-R) was developed (see Appendix B). In the MHC-SF-R, EWB items were maintained, items on the PWB scale were edited for clarity and the SWB scale was restructured into societal well-being (societal WB) and relational well-being (relational WB). Since the SWB scale was reorganized into two new scales, the MHC-SF-R comprises four instead of three dimensions for assessing positive mental health. The revised version contains 19 items, three evaluating EWB, six measuring PWB, six investigating societal WB and four appraising relational WB. In the new societal WB scale, the structure of MHC-SF's old SWB scale which was based on Keyes's (1998) dimensions was taken over. Therefore, the new societal WB scale is similar to the old SWB scale with the amendments that all corresponding items were reframed and a second item measuring social integration was added (see Appendix B). The new relational WB subscale has no direct connection to the old version of the MHC-SF. Prilleltensky (2005) stressed the importance of including relational WB into comprehensive well-being assessments. According to him, relational WB is "reflected in the presence of supportive relationships, which derive from successful experiences of nurturance and attachment, and is promoted by empathy and opportunities to give and receive caring compassion" (Prilleltensky, 2005, p. 56). Since it seems meaningful to incorporate an assessment of relational WB into a comprehensive well-being measurement instrument, a relational WB dimension was integrated in the MHC-SF-R. Moreover, MHC-SF's test instructions and response options were modified. Köhle (2010) found that several survey participants had difficulties remembering how often they had a particular feeling during the last month. Therefore, in the MHC-SF-R test instructions were altered by asking respondents to indicate how often they experienced a certain feeling during the past week instead of asking them how often they experienced a certain feeling during the past month. The response possibilities of the MHC-SF-R were adjusted, e.g. "about once a week" was changed into "sometimes", so that answer possibilities fit the modified test instructions. However, MHC-SF's 6-point Likert-Scale was retained.

The primary aim of this study is to validate for the very first time the MHC-SF-R. For this purpose, its factorial structure will be examined, internal consistency will be established and convergent validity with two questionnaires measuring theoretically related constructs will be investigated. The secondary aim of the current study is to examine how the MHC-SF-R performs in comparison with the MHC-SF in terms of psychometric properties. In this regard, internal consistency and convergent validity of MHC-SF-R's and MHC-SF's total scales and comparable subscales will be contrasted. Since the revised version of the

MHC-SF overcomes many limitations of MHC-SF's original version, it is expected that the MHC-SF-R will show better or similar internal consistency and convergent validity compared with the MHC-SF. For exploring convergent validity, two different psychological constructs which are indicators of individual functioning were chosen, namely self-esteem and self-efficacy (Baumeister, 1993; Gecas, 1989). According to Baumeister (1993), self-esteem is defined as a person's entire sense of worthiness whereat the extent of perceived worthiness depends on how much value a person puts on him- or herself (Baumeister, Campbell, Krueger, & Vohs, 2003). General self-efficacy can be defined as a person's belief in his or her ability to cope with an extensive spectrum of challenging or stressful requirements (Luszczynska, Scholz, & Schwarzer, 2005). The psychological constructs self-esteem and self-efficacy were chosen as validation constructs since previous research revealed that self-esteem and self-efficacy correlate moderately with MHC-SF's total score and especially with its PWB subscale (Keyes et al., 2008; Lamers et al., 2011; Petrillo et al., 2015). Therefore, it can be expected that similar correlations will be found between the MHC-SF-R and the psychological constructs self-esteem and self-efficacy.

In line with the aforementioned aims, the following hypotheses are established. It is hypothesized that, H1: The MHC-SF-R will show an acceptable model fit for a four-factor structure solution (emotional, psychological, societal and relational well-being). H2: At least acceptable internal consistency will be found for MHC-SF-R's total scale and all of its subscales. H3a: At least moderate statistically significant correlations will be found between MHC-SF-R's total score and the total scores of two different questionnaires measuring self-esteem and self-efficacy. H3b: Correlations between MHC-SF-R's PWB subscale score and the total scores of the two different questionnaires measuring self-esteem and self-efficacy will be statistically significant and higher than correlations between MHC-SF-R's other subscale scores with the total scores of the two questionnaires measuring self-esteem and self-efficacy. H4: The MHC-SF-R will show significantly higher or similar internal consistency and convergent validity for all of its scales compared with internal consistency and convergent validity of the corresponding scales belonging to the MHC-SF.

All statistical analysis that will be run for investigating the abovementioned hypothesis formulated with regard to the MHC-SF-R will also be run for the MHC-SF in order to be able to contrast the MHC-SF-R with the MHC-SF in all respects. Besides statistical analysis that are necessary for investigating the hypotheses, intercorrelations between MHC-SF(-R)'s total scales and subscales will be explored and compared.

2 Methods

2.1 Design

In the current study, a questionnaire survey design was implemented. Since all data were collected during a set period of time and no comparison were made between different points in time, the present study is a cross-sectional study. The ethics committee from University of Twente's Faculty of Behavioral, Management and Social Sciences (BMS) approved the study in April 2019. The ethical approval can be requested at the faculty of BMS by means of its registration number 190441.

2.2 Participants

For this study, English speaking students aged 18 years and older who are enrolled at a University or HBO were chosen as the target group. Participants were recruited via personal invitations, Facebook, Instagram and University of Twente's Sona System. University of Twente's Sona System is an online application which conveys research activities implemented by researchers and students of the University of Twente. Additionally, the application keeps track of test subject hours acquired by students who participated as a test subject in research activities. For completing a psychology bachelor's degree at the University of Twente, students have to collect in total 15 test subject hours. Subjects who participated via the Sona System were reimbursed for their involvement with 0.25 subject hours. All other subjects participated without compensation. The way of recruitment implemented in this study implies that mainly students who were easy to reach were approached. Therefore, the current sample is a convenience sample. Exclusion criteria were being younger than 18 years, not being enrolled at a University or HBO and indicating English language proficiency lower than advanced (i.e. beginner or intermediate level).

Altogether, 125 subjects were recruited and started the survey between the 11th of April and the 6th of May 2019. From those subjects, 6 participants were excluded because their English proficiency was not sufficient (i.e. beginner or intermediate level) and 3 other subjects were excluded because they were not enrolled as a student. Since completers only analysis was applied, 7 subjects who did not finish the questionnaire were excluded from

analyses. Additionally, 2 other subjects were excluded because after deleting their datasets data for MHC-SF's PWB subscale and RSE's total scale were normally distributed (see 2.5). Consequently, the final sample appropriate for analysis consisted of 107 participants from which 53 completed the MHC-SF-R and 54 completed the MHC-SF. There was no statistical difference [$t(105) = -1.682$, $p = .095$] in age between subjects who completed the MHC-SF-R ($M = 22.5$, $SD = 3.0$) and those who did the MHC-SF ($M = 21.6$, $SD = 2.3$). Moreover, no statistical difference [$X^2 (1, N = 107) = 0.08$, $p = .777$] was found in gender between participants who completed the MHC-SF-R (52.8% female) and those wo did the MHC-SF (55.5% female). This indicates that randomization was successful.

2.3 Measurement Instruments

2.3.1 Mental Health Continuum - Short Form, revised. The MHC-SF-R is a modified version of the MHC-SF. It is a self-administered questionnaire containing 19 items measuring overall mental well-being. The MHC-SF-R encompasses four dimensions consisting of emotional well-being (3 items), psychological well-being (6 items), societal well-being (6 items) and relational well-being (4 items). Respondents have to indicate on a 6-point Likert scale ranging from 0 "never" to 5 "almost always" how often they experienced a specific feeling during the last week. For interpreting the outcome of the questionnaire, mean scores of MHC-SF-R's total score and subscale scores are calculated by dividing the obtained score on each scale by the number of its corresponding items. The total score (0-95) is calculated by adding up respondents' scores obtained on all items. Subscale scores for EWB (0-15), PWB (0-30), societal WB (0-30) and relational WB (0-20) are calculated by adding up respondents' scores obtained on items belonging to a particular dimension. Higher scores indicate a higher level of mental well-being.

2.3.2 Mental Health Continuum - Short Form. The MHC-SF is a self-administered questionnaire consisting of 14 items measuring overall mental well-being. It comprises three dimensions made up of emotional well-being (3 items), psychological well-being (6 items) and social well-being (5 items). Respondents have to indicate on a 6-point Likert scale ranging from 0 "never" to 5 "every day" how often they experienced a specific feeling during the last month. For interpreting the outcome of the questionnaire, mean scores of MHC-SF's

total score and subscale scores are calculated by dividing the obtained score on each scale by the number of its corresponding items. MHC-SF's total score (0-70) is calculated by adding up respondents' scores on all items. The subscale scores for EWB (0-15), PWB (0-30) and SWB (0-25) are calculated by adding up respondents' scores on items belonging to a specific dimension. Higher scores reflect a higher level of mental well-being.

According to Keyes et al. (2008), Lamers et al. (2011) and Petrillo et al. (2015), MHC-SF's psychometric properties can be considered as adequate to good. MHC-SF's psychometric properties are extensively discussed in the introduction of this paper.

2.3.3 Rosenberg Self-Esteem Scale. The Rosenberg Self-Esteem Scale (RSE) is a self-administered questionnaire measuring global self-esteem (Rosenberg, 1979). It contains 10 items evaluating positive and negative feelings about the self. Five items of the scale are positively phrased and 5 items are negatively phrased. The negative items are reversed scored. Respondents have to rate each item on a 4-point Likert scale ranging from 1 "strongly disagree" to 4 "strongly agree". The total score (10-40) is calculated by adding up respondents' scores obtained on each item, whereat higher scores reflect a higher degree of self-esteem (Rosenberg, 1979).

Previous research demonstrated that the RSE shows a relatively stable one factor structure solution (Gray-Little, Williams, & Hancock, 1997), has good internal consistency (Schmitt & Allik, 2005) and high test-retest reliability (Torrey, Mueser, McHugo, & Drake, 2000). RSE's convergent and discriminant validity can be considered as good (Brumfitt & Sheeran, 1999; McCurdy & Kelly, 1997; Robins, Hendin, & Trzesniewski, 2001). In the current study RSE's internal consistency was high ($\alpha = .85$).

2.3.4 General Self-Efficacy Scale. The General Self-Efficacy Scale (GSE) is a self-administered instrument for measuring general self-efficacy (Schwarzer & Jerusalem, 1995). The English version of the GSE used in this study was translated by Mary Wagner from the original German version (Schwarzer, 1992). The questionnaire comprises 10 items investigating the strength of an individual's conviction in his or her capability to perform well in difficult or novel situations. Respondents have to score each item on a 4-point Likert scale ranging from 1 "Not at all true" to 4 "Exactly true". The total score (10-40) is calculated by adding up respondents' scores obtained on each item, whereat higher scores indicate stronger generalized self-efficacy beliefs (Schwarzer & Jerusalem, 1995).

Previous research indicates that the GSE displays a stable one factor structure solution (Scholz, Doña, Sud, & Schwarzer, 2002; Schwarzer, Bäßler, Kwiatek, Schröder, & Zhang, 1997), has good internal consistency (Scholz et al., 2002) and exhibits moderate test-retest reliability (Schwarzer & Jerusalem, 1999). Convergent validity of the GSE is good (Petrillo et al., 2015; Scholz et al., 2002) and its discriminant validity is adequate (Scholz et al., 2002). In the current study GSE's internal consistency was high ($\alpha = .86$).

2.4 Procedure

The survey was administered online via Qualtrics between the 11th of April and the 6th of May 2019. Participants who registered for the study were provided with a link. By means of the link, they were able to start the online questionnaire. At the start of the questionnaire, subjects were welcomed and informed about the purpose and the duration of the study. They were explained that a newly developed questionnaire for assessing overall mental well-being needs to be validated and that the duration for completing the survey takes approximately 15-20 minutes. Further, subjects were asked to answer all items to the best of their knowledge. After providing participants with general information about the study, online informed consent was obtained (see Appendix C). After providing informed consent, participants were screened for eligibility and asked for their demographics (i.e. age, gender). Provided that subjects meet all inclusion criteria, they were randomly assigned and redirected to either the MHC-SF-R or the MHC-SF. Consequently, two independent groups were established. This was done because it was assumed that response bias might occur if participants fill in the MHC-SF-R and the MHC-SF in a row. After filling in either the MHC-SF-R or the MHC-SF all subjects were asked to complete the GSE and the RSE. In the end of the study, subjects were again provided with contact information for approaching the research term for asking questions or getting informed about the outcome of the study.

2.5 Analysis

For analyzing data SPSS 24.0, LISREL 9.3 and R 3.4.4 were utilized. SPSS 24.0 was used for determining outliers concerning all scales, checking baseline differences between groups (MHC-SF-R, MHC-SF) regarding participant's demographics (age, gender), establishing internal consistency for all measurement instruments including their corresponding subscales and determining convergent validity for both versions of the MHC-SF and their subscales with the RSE and GSE. For all analyses, completers-only analyses were implemented and where applicable, the significance level α was set to .05.

For determining outliers, the outlier labeling rule as defined by Hoaglin and Iglewicz (1987) was employed by multiplying the interquartile range of all scales used in this study with the factor 2.2 for establishing their lower and upper bounds. In doing so, two outliers were detected. One subject exceeded the lower bound for MHC-SF's PWB scale. The other exceeded the lower bound of RSE's total scale. Both subjects were removed from analysis because after deleting their datasets, data for MHC-SF's PWB subscale and RSE's total scale became normally distributed.

Before analyzing the dataset, different normality tests were conducted. The normality test Shapiro-Wilk indicated for most of the scales that data is not normally distributed [MHC-SF-R's total scale (p = .027), MHC-SF-R's EWB scale (p = .006), MHC-SF-R's societal WB scale (p = .018), MHC-SF-R's relational WB scale (p = .010) and MHC-SF's total scale (p = .007), MHC-SF's EWB scale (p = .003), MHC-SF's SWB scale (p = .023)] whereas a less strict normality test, suggested by Kim (2013), pointed out that all scales are normally distributed. Due to the contradictory normality test results, additionally histograms with normality curves and Q-Q plots were examined. The overall picture of the three different approaches for checking normality indicated that data for all scales is at least approximately normally distributed. Detailed information about the outcomes of the Shapiro-Wilk test, the normality test suggested by Kim (2013) and the eyeball test can be found in Appendix D.

For the purpose of checking baseline differences between subjects who completed the MHC-SF-R and those who did the MHC-SF in age, an independent t-test was run with "group" (MHC-SF-R, MHC-SF) as independent variable and "age" as dependent variable. Baseline differences in gender between groups were checked by means of a chi-square test of independence.

For examining MHC-SF-R's and MHC-SF's model fits, LISREL 9.3 was used for conducting confirmatory factor analysis. Since normality tests performed in LISREL indicated that data for the MHC-SF-R and MHC-SF is not normally distributed but normality tests which are less strict indicated that data is normally distributed (Kim, 2013), robust maximum likelihood is the appropriate estimation method (Li, 2016). However, the sample sizes in both groups were too small for calculating asymptotic covariance matrices which are necessary for conducting robust maximum likelihood estimation method. For this reason, normal maximum likelihood estimation (Chou & Bentler, 1995) was chosen as estimation method. All items were supposed to load on one factor and the variance of factors was fixed to 1. For interpreting factor loadings, standardized coefficients were investigated. According to Brown (2015), factor loadings in standardized solutions can be interpreted as the correlation of the item with the latent factor if the model has no double-loading items on latent factors. Therefore, squaring a standardized factor loading gives the proportion of variance for the item that is explained by the factor. According to Hair, Black, Babin, Anderson, and Tatham (2006), factor loadings (≤ .33) translating to less than 10 % explanation are seen as problematic, factor loadings (.34 - .7) indicating 11 % and 49 % explanation by the factor are seen as acceptable and factor loadings (> .7) exceeding 50 % of variance explained by the factor are considered as high. For assessing model fit, different fit indices were used: Chi-Square (X^2) significance test, Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI) and Comparative Fit Index (CFI). For determining if the model displayed acceptable model fit, cut-off values recommended from Hu and Bentler (1999) were applied. RMSEA (for sample sizes $N < 250$) and SRMR values between .08 and .10 indicate acceptable model fit, values < .08 good model fit and values > .10 poor model fit. For GFI, AGFI and CFI, a value ≥ .90 suggests acceptable model fit and a value > .95 good model fit. Values < .90 indicate poor model fit.

Internal consistency was established by means of Cronbach's alpha (Cronbach, 1951). Following Kline (2000), Cronbach's alpha values ≥ .7 are interpreted as acceptable and ≥ .8 as high. Cronbach's alpha values < .7 are referred to as inacceptable. For investigating differences in internal consistencies between groups (MHC-SF-R/ MHC-SF) the R package "Cocron" was run in R 3.4.4. The R package "Cocron" (COmparing CRONbach's alphas) can be used for comparing two or more alpha values established from either independent or dependent groups (Diedenhofen & Musch, 2016). The implementation of comparing

Cronbach's alphas is based on methods illustrated by Feldt, Woodruff, and Salih (1987). The test statistic for contrasting *m* independent alpha coefficients rests on a transformation of the alpha coefficients (Diedenhofen & Musch, 2016). During transformation, the number of participants per group and the number of items in each test is taken into consideration. If the null hypothesis is true, the test statistic is roughly distributed as X^2 with $df = m\text{-}1$ (Diedenhofen & Musch, 2016). For comparing Cronbach's alphas between MHC-SF-R's and MHC-SF's total and subscales, the independent variable was "group" (MHC-SF-R/ MHC-SF) and the dependent variable "Cronbach's alpha".

Convergent validity was established by correlating MHC-SF-R's and MHC-SF's total and subscale scores with RSE and GSE total scores. For calculating correlations, Pearson's correlation coefficient was chosen since all scales are treated as continuous data, significant outliers were excluded, data is approximately normal distributed and scatter plots revealed that there is a linear relationship between correlated variables. Two tailed significance testing was conducted because it could not be accurately predicted that statistical significance goes only in one direction. When performing correlation analysis, the correlation coefficient can be interpreted as the effect size. Therefore, the correlation coefficient can be used to make inferences about the strength of the association between the correlated constructs (Gall, Gall, & Borg, 2007). In the current paper, correlation values for convergent validity were interpreted in accordance with Cohen (1992). Therefore, correlation values < .1 are considered as weak, correlations between .1 and .3 as small, correlations between .3 and .5 as moderate and correlations ≥ .5 as high. For investigating differences in convergent validity between groups (MHC-SF-R/ MHC-SF) the R package "Cocor" was run in R 3.4.4. The R package "Cocor" (COmparing CORalations) can be used for comparing two correlation coefficients established from either independent or dependent groups with overlapping or nonoverlapping variables (Diedenhofen & Musch, 2015). The implementation of comparing two correlation coefficients established from two independent groups of different sample sizes is based on Fisher's *r* to *z* transformation (Fisher, 1925). After both correlation coefficients are transformed into *z*-scores, the *z*-scores are analyzed for statistical significance by calculating the *Z*-test statistic by means of the following formula: $Z = (z_1 - z_2) / \text{square root of } [(1 / N_1 - 3) + (1 / N_2 - 3)]$. When setting $\alpha = .05$, the null hypothesis is assumed to be true if the *Z* statistic is between ± 1.96 (Diedenhofen & Musch, 2015). For the same reason as mentioned above, two tailed significance testing was conducted. When comparing Person's correlation coefficients that were established between the MHC-SF-R and the validation

questionnaires (RSE/ GSE) with the corresponding correlation coefficients that were identified between the MHC-SF and the validation questionnaires, the independent variable was “group” (MHC-SF-R/ MHC-SF) and the dependent variable “correlation coefficient”.

3 Results

3.1 Factor Structure

The model fit indices for MHC-SF-R’s four factor solution (EWB, PWB, societal WB, relational WB) and MHC-SF’s three factor solution (EWB, PWB, SWB) are shown in table 1. All fit indices suggest that MHC-SF-R’s model fit and MHC-SF’s model fit is not acceptable, except the fit index SRMR for MHC-SF’s three factor solution (SRMR = .100).

Table 1

Maximum Likelihood Estimation of CFA Models of the Latent Structure of the MHC-SF-R Items and of the MHC-SF Items

Fit indices	MHC-SF-R’s Four Factor Solution (EWB, PWB, societal WB, relational WB) (N = 53)	MHC-SF’s Three Factor Solution (EWB, PWB, SWB) (N = 54)
X^2	226.366***	120.002***
df	146	74
RMSEA	.102	.107
SRMR	.109	.100
GFI	.691	.763
AGFI	.598	.663
CFI	.820	.833

Note. CFA = confirmatory factor analysis; MHC-SF-R = Mental Health Continuum Short Form revised; MHC-SF = Mental Health Continuum Short Form; X^2 = chi-square; *df* = degrees of freedom; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; GFI = Goodness of Fit Index; AGFI = Adjusted Goodness of Fit Index; CFI = Comparative Fit Index; CFA = Confirmatory Factor Analysis; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; WB = well-being; *** = $p < .001$.

Figure 1 shows the standardized coefficients established from MHC-SF-R's four factor solution. All standardized factor loadings for items loading on the EWB scale were high (.76 - .94). For the PWB scale, question 13 displayed a problematic factor loading whereas all other items loading on the PWB scale were acceptable (.51 - .69). Items loading on the societal WB scale showed the biggest range of factor loadings (.17 - .92). For the societal WB scale, question 6 and question 8 displayed problematic factor loadings while question 19 indicated a high factor loading. All other items loading on the societal WB scale exhibited acceptable factor loadings. With respect to the relational WB scale, factor loadings ranged from (.55 - .86) indicating acceptable to high factor loadings.

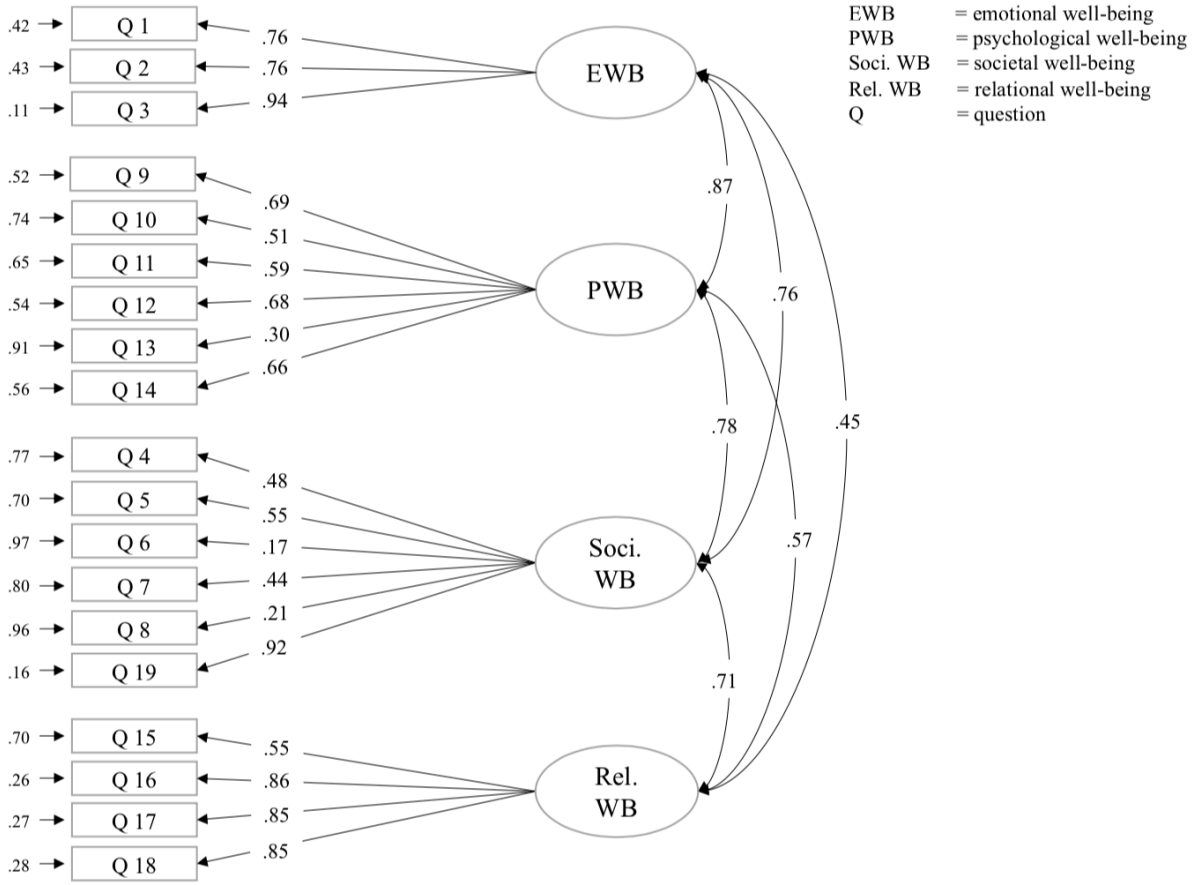

Figure 1. Confirmatory factor analysis of MHC-SF-R's four factor solution with standardized coefficients.

Figure 2 displays standardized coefficients established from MHC-SF's three factor solution. For all three subscales of the MHC-SF acceptable to high standardized factor loadings were found [EWB: (.54 - .81); PWB: (.51 - .71); SWB (.61 - .82)].

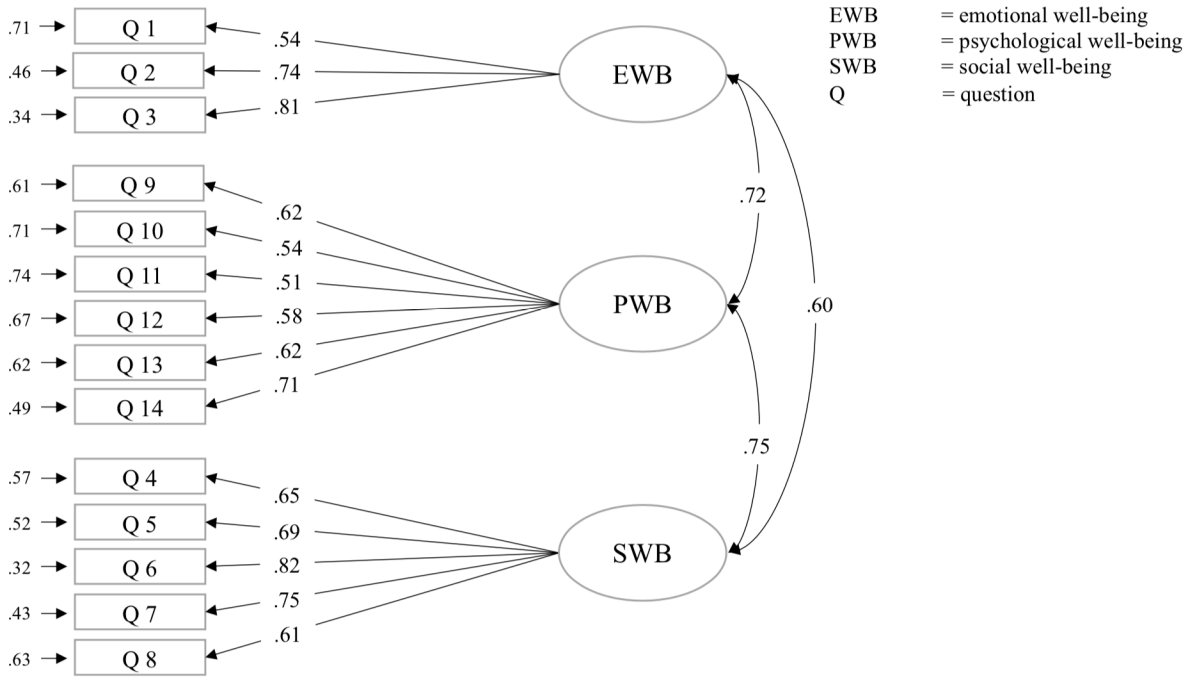

Figure 2. Confirmatory factor analysis of MHC-SF's three factor solution with standardized coefficients.

3.2 Means and Intercorrelations between MHC-SF-(R)'s Total and Subscale Scores

Table 2 and table 3 show the mean scores, standard deviations and correlational statistics sustained from MHC-SF-R's total score and its subscales and MHC-SF's total score and its subscales, respectively. For MHC-SF-R's subscales, the lowest mean score was found for its societal WB scale (M = 3.2, SD = 0.7) and the highest for its PWB scale (M = 3.8, SD = 0.7). Regarding subscales belonging to the MHC-SF, the lowest mean score was observed for its SWB scale (M = 2.6, SD = 1.0) and the highest for its PWB scale (M = 3.8, SD = 0.7).

For the MHC-SF-R and the MHC-SF, all intercorrelations were significant ($p < .01$) and positive indicating that subjects who scored higher on any dimension of the MHC-SF(-R) scored also higher on the other dimensions. For the MHC-SF-R and MHC-SF, intercorrelations established between their total scales and corresponding subscales were higher than intercorrelations within their subscales. With regard to the MHC-SF-R and its

subscales, the lowest intercorrelation was found between its societal WB scale and its EWB scale (r = .36) and the highest intercorrelation was detected between its total scale and the PWB scale (r = .85). Regarding the MHC-SF and its subscales, the lowest intercorrelation was found between its SWB and PWB scale (r = .44) and the highest intercorrelation was established between its total scale and its PWB scale (r = .89). Therefore, intercorrelations for the MHC-SF-R and the MHC-SF are generally in the same range. For all established intercorrelations, the strength of the relationships is moderate to high.

Table 2

Pearson's Correlation Coefficients of MHC-SF-R's Subscale Mean Scores and Total Mean Score (N = 53)

	M (SD)	1	2	3	4	5
Total MHC-SF-R	3.5 (0.6)	1				
EWB	3.5 (1.0)	.75**	1			
PWB	3.8 (0.7)	.85**	.69**	1		
Societal WB	3.2 (0.7)	.79**	.36**	.49**	1	
Relational WB	3.7 (0.9)	.80**	.43**	.51**	.60**	1

Note. All p values are two-tailed. MHC-SF-R = Mental Health Continuum Short Form Revised; EWB = emotional well-being; PWB = psychological well-being; WB = well-being; ** = $p < .01$.

Table 3

Pearson's Correlation Coefficients of MHC-SF's Subscale Mean Scores and Total Mean Score (N = 54)

	M (SD)	1	2	3	4
Total MHC-SF	3.2 (0.8)	1			
EWB	3.5 (0.8)	.68**	1		
PWB	3.5 (0.8)	.89**	.51**	1	
SWB	2.6 (1.0)	.88**	.44**	.61**	1

Note. All p values are two-tailed. MHC-SF = Mental Health Continuum Short Form; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; ** = $p < .01$.

3.3 Internal Consistency

In table 4, internal consistency is given for the MHC-SF-R, the MHC-SF and their corresponding subscales. For MHC-SF-R's total scale, its EWB scale and relational WB scale internal consistency was high. MHC-SF-R's PWB scale displayed acceptable internal consistency while MHC-SF's societal WB scale demonstrated moderate internal consistency. MHC-SF's total scale and SWB scale showed high internal consistency whereas its EWB scale and PWB scale demonstrated acceptable internal consistency.

Table 4

Internal Consistency (Cronbach's Alphas) for the MHC-SF-R, the MHC-SF and Their Corresponding Subscales Including Chi-Square Statistics for Testing Statistical Significance Between Cronbach's Alphas

	MHC-SF-R (N = 53)	MHC-SF (N = 54)	X^2	p
Total Scale	.89	.88	0.04	.820
EWB	.88	.75	4.88	.027*
PWB	.75	.77	0.06	.800
Societal WB and SWB	.65	.83	4.54	.033*
Relational WB	.85	-	-	-

Note. All p-values are two-tailed. Cronbach's alpha for the societal WB scale is given for the MHC-SF-R and Cronbach's alpha for the SWB scale for the MHC-SF. MHC-SF-R = Mental Health Continuum Short Form Revised; MHC-SF = Mental Health Continuum Short Form; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; WB= well-being; X^2 = Chi-Square statistics calculated according to methods suggested by Feldt et al. (1987) with (df = 1); * = $p < 0.05$.

Significant differences between Cronbach's alphas were found between MHC-SF-R's and MHC-SF's EWB scales [X^2 (1, N = 107) = 4.88, p = .027], indicating that Cronbach's alpha of MHC-SF-R's EWB scale (α = .88) is significantly higher than Cronbach's alpha of MHC-SF's EWB scale (α = .75). Moreover, there was a significant difference between MHC-SF-R's societal and MHC-SF's SWB dimension [X^2 (1, N = 107) = 4.54, p = .033], demonstrating that Cronbach's alpha of MHC-SF-R's societal dimension (α = .65) is significantly lower than Cronbach's alpha of MHC-SF's SWB dimension (α = .83). For the PWB dimensions and the total scales no significant differences between Cronbach's alphas were found.

3.4 Convergent Validity

In table 5, Pearson's correlation coefficients are given for MHC-SF-R's and MHC-SF's calculated associations with the RSE. Additionally, Fisher's *Z*-test statistics are displayed resulting from comparing the correlation coefficients that were established between the MHC-SF-R and the RSE with the corresponding correlation coefficients that were identified between the MHC-SF and the RSE. In table 6, Pearson's correlation coefficients for MHC-SF-R's and MHC-SF's associations with the GSE including Fisher's *Z*-test statistics for contrasting the coefficients are illustrated. All calculated correlation coefficients were positive indicating that all correlated variables move in the same direction. Moreover, all determined correlation coefficients were statistically significant ($p < .05$), except the correlation coefficient established between the GSE and MHC-SF-R's societal WB scale and the correlation coefficient calculated between the GSE and MHC-SF's SWB scale. With regard to the RSE, this means that subjects who scored higher on self-esteem also scored higher on all evaluated dimensions of mental well-being. Concerning the GSE, this indicates that subjects who scored higher on self-efficacy also scored higher on emotional, psychological, relational and overall mental wellbeing, but that there is no association between self-efficacy scores and societal WB/ SWB scores. All significant correlation coefficients indicated moderate to high associations.

Correlation coefficients calculated between MHC-SF-R's scales and the RSE were highest for MHC-SF-R's PWB subscale ($r = .68$) followed by its total scale ($r = .62$). Regarding correlation coefficients calculated between MHC-SF-R's scales and the GSE, correlation coefficients were highest for MHC-SF-R's EWB subscale ($r = .58$) followed by its total scale ($r = .47$).

For the MHC-SF, correlation coefficients established between the RSE and MHC-SF's total scale were highest ($r = .57$) followed by its PWB scale ($r = .55$). Furthermore, correlation coefficients determined between the GSE and the MHC-SF were highest for MHC-SF's EWB scale ($r = .43$) followed by its PWB scale ($r = .42$). However, correlations coefficients established between the GSE and MHC-SF were nearly the same for its total scale ($r = .41$), EWB scale ($r = .43$) and PWB scale ($r = .42$).

When comparing the correlation coefficients established between the MHC-SF-R and the RSE with the correlation coefficients calculated between MHC-SF's corresponding scale and the RSE, not any statistical difference between the correlation coefficients was found, see table 5. The same applies when comparing the correlation coefficients determined between

the MHC-SF-R and the GSE with the coefficients identified between MHC-SF's corresponding scales and the GSE, see table 6.

Table 5

Pearson's Correlation Coefficients Indicating the Associations Between Rosenberg's Self-Esteem Scale and Both Versions of the MHC-SF with Subscales Including Fisher's Z-Test Statistic with Corresponding p-Value for Testing Statistical Significance Between Correlation Coefficients Established From Two Different Groups

	Rosenberg's Self-Esteem Scale ($N = 107$)		Fisher's Z	p
	MHC-SF-R ($N = 53$)	MHC-SF ($N = 54$)		
MHC-SF(-R) Total Scale	.62**	.57**	.39	.696
EWB	.57**	.43**	.94	.344
PWB	.68**	.55**	1.06	.288
Societal WB and SWB	.38**	.43**	-.30	.762
Relational WB	.34*	--	--	--

Note. All *p*-values are two-tailed. The societal WB correlation coefficient is given for the MHC-SF-R and the SWB correlation coefficient for the MHC-SF. MHC-SF-R = Mental Health Continuum Short Form Revised; MHC-SF = Mental Health Continuum Short Form; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; WB = well-being; Fisher's $Z = Z$ statistics for comparing correlation coefficients calculated according to Fisher (1925); -- = no value computable; * = $p < 0.05$; ** = $p \leq 0.01$.

Table 6

Pearson's Correlation Coefficients Indicating the Associations Between the General Self-Efficacy Scale and Both Versions of the MHC-SF with Subscales Including Fisher's Z-Test Statistic with Corresponding p-Value for Testing Statistical Significance Between Correlation Coefficients Established From Two Different Groups

	General Self-Efficacy Scale ($N = 107$)		Fisher's Z	p
	MHC-SF-R ($N = 53$)	MHC-SF ($N = 54$)		
MHC-SF(-R) Total Scale	.47**	.41**	.37	.708
EWB	.58**	.43**	1.01	.308
PWB	.42**	.42**	0	1.00
Societal and SWB	.15	.24	-.47	.638
Relational WB	.41**	--	--	--

Note. All *p*-values are two-tailed. The societal WB correlation coefficient is given for the MHC-SF-R and the SWB correlation coefficient for the MHC-SF. MHC-SF-R = Mental Health Continuum Short Form Revised; MHC-SF = Mental Health Continuum Short Form; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; WB = well-being; Fisher's $Z = Z$ statistics for comparing correlation coefficients calculated according to Fisher (1925); -- = no value computable; ** = $p \leq 0.01$.

4 Discussion and Conclusion

4.1 General Discussion

The primary aim of the present study was to validate for the very first time the MHC-SF-R, a self-report questionnaire for assessing positive mental health. Against expectations, all model fit indices for MHC-SF-R's four factor solution indicated that its model fit is not acceptable suggesting that its factorial validity is poor. Therefore, H1 is rejected. As expected, for MHC-SF-R's EWB, PWB, relational WB and total scale internal consistency was at least acceptable, meaning that H2 is accepted for those scales. Contrary to expectations, internal consistency for MHC-SF-R's societal WB scale was not acceptable, which means that H2 is rejected for the societal WB scale. As presumed, the correlations established between MHC-SF-R's total scale and RSE's total scale and between MHC-SF-R's total scale and GSE's total scale were significant and high, meaning that H3a is accepted. Moreover, the correlation between MHC-SF-R's PWB subscale and RSE's total scale was significant and higher than the correlations determined between MHC-SF-R's other subscales with RSE's total scale, indicating that H3b is accepted with regard to self-esteem. Against expectations, the correlation between MHC-SF-R's PWB subscale and GSE's total scale was not higher than the correlation between MHC-SF-R's EWB subscale with GSE's total scale, which means that H3b is rejected with respect to self-efficacy.

The secondary aim of the current study was to examine how the MHC-SF-R performs in comparison with the MHC-SF in terms of psychometric properties. Concerning internal consistency, as expected, Cronbach's alphas for MHC-SF-R's EWB, PWB and total scale were either significantly higher or not statistically different compared with Cronbach's alphas for MHC-SF's corresponding scales. Against expectations, MHC-SF-R's societal WB scale performed significantly worse in terms of internal consistency compared with MHC-SF's SWB scale. Regarding convergent validity, as hypothesized, not any statistical significant difference was found when comparing correlation coefficients that were established between the MHC-SF-R and the RSE/ GSE with the corresponding correlation coefficients that were identified between the MHC-SF and the RSE/ GSE. Therefore, with regard to convergent validity, H4 is accepted for the comparison of all scales. Concerning internal consistency, H4 is accepted for the comparison of all scales, except for the comparison between MHC-SF-R's societal WB and MHC-SF's SWB dimension.

At first glance, it seems that the MHC-SF-R lacks factorial validity whilst showing high internal consistency and high convergent validity for its total scale. In the next sections all main findings will be discussed in depth. In doing so, findings on MHC-SF-R's factorial structure and psychometric properties will be juxtaposed with findings on MHC-SF's factorial structure and psychometric properties. Moreover, strengths and limitations of the present study will be elucidated and implications for practice and future research are given.

4.1.1 Factor Structure. For determining MHC-SF-R's model fit, a wide assortment of fit indices were checked because there is not a single measure for determining the fit of a model when performing confirmatory factor analysis (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Unexpectedly, all established model fit indices for MHC-SF-R's four factor solution indicated that its model fit is not acceptable. However, fit indices established for the MHC-SF also suggest that its three factor solution is not acceptable. This finding is very surprising since previous research repeatedly showed that the MHC-SF shows a stable three factor solution (Echeverría et al., 2017; Franken et al., 2018; Keyes, 1998; Lamers et al., 2011; Perugini et al., 2017; Petrillo et al., 2015).

As the two models are composed of different factor structures and contain partly different items, their fit indices cannot be compared directly. However, when contrasting the overall picture of MHC-SF-R's fit indices with the overall appearance of MHC-SF's fit indices, the three factor structure of the MHC-SF seems to fit the model slightly better than the four factor structure of the MHC-SF-R because the fit indices GFI, AGFI, CFI and SRMR get closer to values that indicate an acceptable model fit (Hu & Bentler, 1999). Moreover, it is striking that for the MHC-SF all standardized factor loadings are acceptable to high, whereas for the MHC-SF-R three items displayed problematic factor loadings, namely question 13 loading on MHC-SF-R's PWB scale, and question 6 and 8 loading on MHC-SF-R's societal WB scale. Since the proportion of variance that is explained by the factor they load on is for the three items less than 10 %, it should be considered to use the corresponding unrevised items (see Appendix B) because the proportion of variance that is explained by the factor they load on for each individual corresponding unrevised item is more than 37 %.

It is notable, that all problematic factor loadings were found for items belonging to the MHC-SF-R. All those items (question 6, 8 and 13) are items that were taken over from the MHC-SF and edited for clarity before incorporating them in the MHC-SF-R. However, the low factor loadings are an indicator that editing those items might have made them more ambiguous instead of less ambiguous. Moreover, it is striking that the societal WB scale

displayed a huge range of factor loadings (.17 -.92) from which only one factor loading was >.55, namely the factor loading of question 19. This indicates that question 19 made up mostly of MHC-SF-R's societal WB dimension.

Additionally, it is remarkable that three out of four items loading on MHC-SF-R's newly generated relational WB scale displayed high factor loadings and the remaining item loading on the relational WB scale showed an acceptable factor loading, suggesting that the new added variables load appropriately on MHC-SF-R's fourth dimension.

4.1.2 Intercorrelations between MHC-SF(-R)'s Total and Subscale Scores. All intercorrelations calculated between MHC-SF-R's total scale and its subscales were higher than intercorrelations established between MHC-SF-R's subscales. This indicates that MHC-SF-R's overall score reflects adequately the outcomes of its individual subscales. In accordance with previous research (Keyes et al., 2008; Lamers et al., 2011; Perugini et al., 2017; Petrillo et al., 2015) the same pattern of intercorrelations was found for MHC-SF's total scale and subscales.

Intercorrelations for MHC-SF-R's subscales and intercorrelations for MHC-SF's subscales were moderate to high, indicating that all subscales within the MHC-SF-R and within the MHC-SF are moderately to highly associated. This suggests that the subscales measure overlapping yet independent constructs of mental well-being.

For the MHC-SF-R and the MHC-SF intercorrelations between their total scales with their PWB scales where highest which is also consistent with previous findings (Keyes et al., 2008; Lamers et al., 2011; Petrillo et al., 2015). Moreover, for the MHC-SF-R and MHC-SF intercorrelations between their EWB scales and societal WB/ SWB scales were lowest which is also congruent with findings from Lamers et al. (2011), Petrillo et al. (2015) and Perugini et al. (2017).

4.1.3 Internal Consistency. For all of MHC-SF-R's scales, except its societal WB scale, internal consistency was acceptable or high. Cronbach's alphas for MHC-SF-R's total scale and MHC-SF's total scale were nearly identical indicating that taken the measurement instruments as a whole, the MHC-SF-R performs in terms of internal consistency approximately equally well as the MHC-SF.

Even though H4 states that it was expected that the MHC-SF-R will show significantly higher or similar internal consistency for all its scales compared with internal consistency of the corresponding scales belonging to the MHC-SF, the finding that

Cronbach's alpha for MHC-SF-R's EWB scale was significantly higher than Cronbach's alpha for MHC-SF's EWB scale is somewhat surprising. For this particular dimension it was rather expected that Cronbach's alphas do not differ statistically since EWB items are the same in the MHC-SF-R and the MHC-SF. However, the modified test instructions and response format (in the MHC-SF-R subjects had to indicate how they felt within the last week instead of how they felt within the last month) might be an explanation why Cronbach's alpha was higher for MHC-SF-R's EWB scale since people are better in accurately remembering what they experienced recently than in what they experienced in a prolonged time (Murre & Dros, 2015).

It is striking that MHC-SF-R's societal WB scale performed bad compared with all other internal consistency estimates established in the current study. Moreover, the societal WB scale performed significantly worse than the SWB scale in terms of internal consistency. Since items belonging to the societal WB scale were revised for being less ambiguous, it was expected that the societal WB scale performs significantly better or at least similar compared with the SWB scale. However, the finding that the societal WB scale performed significantly worse than the SWB scale is a further indication that at least some revisions of SWB items might have made the societal WB scale less stable compared to its original version. Another more extensive but speculative explanation for the finding that the revised SWB items (i.e. the new societal WB items) did not lead to acceptable internal consistency might be that the underlying structure of the societal WB and SWB dimension, which is based on Keyes (1998), might cause difficulties. In favor of this assumption is that previous research showed that MHC-SF's SWB dimension shows problematic internal consistency (Keyes et al., 2008) and that revising the SWB items did not lead to acceptable internal consistency in the present study. Against this assumption is that other researchers found adequate internal consistency for MHC-SF's SWB scale (Lamers et al., 2011; Petrillo et al., 2015) and that in the current study internal consistency for MHC-SF's SWB scale was high.

With regard to the relational WB scale, internal consistency for the new introduced scale was high suggesting that the scale is reliable. Moreover, internal consistency of MHC-SF-R's total scale including the relational WB items was high. Therefore, is seems meaningful to integrate the relational WB scale in a measurement instrument which investigates overall mental well-being.

4.1.4 Convergent Validity. For examining convergent validity for MHC-SF-R's total scale and subscales, two different psychological constructs which are indicators of individual functioning were chosen, namely self-esteem and self-efficacy (Baumeister, 1993; Gecas, 1989). Self-esteem was measured by means of the RSE and self-efficacy with the aid of the GSE.

The correlation coefficient established between MHC-SF-R's total scale and the RSE was high, indicating that convergent validity between the MHC-SF-R and the psychological construct self-esteem is high. Furthermore, the correlation coefficient found between MHC-SF-R's total scale and the GSE was moderate, suggesting that convergent validity between the MHC-SF-R and the psychological construct self-efficacy is moderate. The correlation coefficients identified between MHC-SF-R's total scale and the RSE and between MHC-SF-R's total scale and the GSE both indicate that the correlated questionnaires measure overlapping yet independent psychological constructs. With regard to the MHC-SF, the identified correlation coefficient between MHC-SF's total scale and the RSE was also high which is in accordance with previous findings from Petrillo et al. (2015). The correlation coefficient determined between MHC-SF's total scale and the GSE was moderate which is as well in line with previous findings (Keyes et al., 2008; Petrillo et al., 2015).

Since the RSE, the GSE and MHC-SF's PWB scale measure components of individual functioning (Baumeister, 1993; Gecas, 1989; Ryff, 1989) it was predicted that MHC-SF-R's PWB scale correlates stronger with the RSE and the GSE than MHC-SF-R's other subscales with the RSE and GSE. As assumed, the correlation between MHC-SF-R's PWB scale and the RSE was higher than correlations between MHC-SF-R's other subscales with the RSE suggesting that convergent validity between MHC-SF-R's PWB subscale and the psychological construct self-esteem is especially high. Regarding the MHC-SF, the correlation established between MHC-SF's PWB subscale and the RSE was highest as well, followed by the correlation determined between MHC-SF's EWB subscale with the RSE. Those findings are in accordance with previous research. Petrillo et al. (2015) also found that MHC-SF's PWB subscale is strongest associated with the RSE followed by the association between MHC-SF's EWB subscale with the RSE. However, as the subscales of the MHC-SF(-R) are all moderately to highly correlated, the correlation coefficients found between MHC-SF(-R)'s EWB subscale with the RSE and MHC-SF(-R)'s PWB subscale with the RSE are relatively close to each other.

Inconsistent with expectations, the correlation coefficient calculated between MHC-SF-R's PWB subscale and the GSE was not the highest correlation coefficient found between

MHC-SF-R's subscales and the GSE, yet it was the second highest. The highest correlation coefficient was established between MHC-SF-R's EWB subscale with the GSE. Nevertheless, this finding suggests that convergent validity between MHC-SF-R's PWB scale and the GSE is moderate. With regard to the MHC-SF, MHC-SF's PWB subscale and MHC-SF's EWB subscale were nearly identically associated with the GSE. These findings are very surprising and contradictory to findings from previous research since Keyes et al. (2008) and Petrillo et al. (2015) both showed that MHC-SF's PWB subscale is considerably strongest correlated with the GSE in comparison with other subscales of the MHC-SF. However, the fact that intercorrelations between MHC-SF(-R)'s subscales are moderate to high might be an explanation why findings in the current study were different.

When comparing the correlation coefficients that were established between the MHC-SF-R and the RSE with the corresponding correlation coefficients that were found between the MHC-SF and the RSE, no significant difference was found for any scale, neither for the comparison of the total scales nor for the comparison of the subscales. The same is true for the comparison of coefficients that were found between the MHC-SF-R and the GSE and the coefficients that were established between the MHC-SF and the GSE. These findings suggest that the MHC-SF-R and the MHC-SF perform equally well in terms of convergent validity with regard to the psychological constructs self-esteem and self-efficacy.

4.2 Strengths and Limitations

The current study comprises unique strengths but also limitations. Due to the inclusion of two groups (MHC-SF-R, MHC-SF) the present study is the first one which investigated MHC-SF-R's psychometric properties and simultaneously offers the possibility to compare those findings with outcomes of MHC-SF's psychometric properties. Since half of the sample completed the MHC-SF-R and approximately the other half of the sample was allocated to the MHC-SF, it was controlled for response bias which might have occurred if all participants would have completed the MHC-SF-R and the MHC-SF (Mortel, 2008).

However, the aforementioned strength of the present study is simultaneously a limitation. Due to splitting the sample into two groups (MHC-SF-R, MHC-SF) the current study design did not allow to check for equivalence between the MHC-SF-R and the MHC-SF through calculating correlations between MHC-SF-R's and MHC-SF's total scores

or subscale scores. Moreover, the comparison of internal consistency and convergent validity between the MHC-SF-R and the MHC-SF relies on the comparison of Cronbach's alphas and correlation coefficients determined from two different groups. Therefore, the comparison of internal consistency and convergent validity between the MHC-SF-R and the MHC-SF is not as valid as it would have been if all subjects would have filled in both versions of the MHC-SF. Nonetheless, this limitation cannot be overcome when the aim is to control for response biases which might occur if the same subjects fill in the MHC-SF-R and the MHC-SF in a row.

A major limitation of the current study concerns the implementation of confirmatory factor analysis for assessing MHC-SF-R's and MHC-SF's model fits. The current sample size was too small for applying the appropriate estimation method for doing confirmatory factor analysis. Since, data for the MHC-SF-R and MHC-SF was approximately normally distributed, which implies that assumptions for normality are slightly violated, robust maximum likelihood estimation would have been the estimation method of choice (Li, 2016; Satorra & Bentler, 1994). However, due to the relatively small sample size it was not possible to calculate the asymptotic covariance matrices which are necessary for implementing robust maximum likelihood estimation (Satorra & Bentler, 1994). Therefore, normal maximum likelihood estimation method was applied which is normally used when data is normally distributed (Chou & Bentler, 1995). This means that the used estimation method did not completely fit the data. Apart from this, the sample size was generally small for conducting confirmatory factor analysis. Some studies suggest that sample sizes (per group) should be larger than 100 (Anderson & Gerbing, 1988; Ding, Velicer, & Harlow, 1995) whereas other authors argue that the appropriateness of a sample size is related to the ratio of subjects to free parameters (Bentler & Chou, 1987) which should be between 5:1 and 10:1 when applying maximum likelihood estimation. However, all recommendations have in common that they indicate that the current sample size might have been too small for getting reliable results. This could be an explanation why contradictory with previous research, model fit indices for the MHC-SF were not acceptable. Moreover, this limitation and its recognition has a major influence on the interpretation of MHC-SF-R's established fit indices. Since the sample size was too small, an inappropriate estimation method was used and contradictory to previous research MHC-SF's fit indices suggest inacceptable model fit for the MHC-SF, it can reasonably be questioned if MHC-SF-R's model fit is indeed inacceptable or if the current finding is misleading.

The next limitation concerns the comparison between Cronbach's alphas of MHC-SF-R's total scale and MHC-SF's total scale. Since the MHC-SF-R and the MHC-SF investigate in part different dimensions of mental well-being (both investigate EWB, PWB and societal WB/ SWB but the MHC-SF-R examines additionally relational WB) a direct comparison of those Cronbach's alphas has to be interpreted with caution. However, since Cronbach's alphas for MHC-SF-R's total scale and MHC-SF's total scale were both very high and very close to each other, it can be reasonably concluded that both total scales perform approximately equally well.

4.3 Implications for Practice and Future Research

Despite the above mentioned limitations, the current study provides valuable implications for practice and future research. MHC-SF-R's and MHC-SF's model fit indices established by means of confirmatory factor analysis suggest that their model fits are not acceptable. However, since the MHC-SF-R and the MHC-SF performed approximately equally in terms of fit indices, it can be assumed that their factorial validity is roughly the same. As previous research has repeatedly shown that MHC-SF's three factor solution displays acceptable model fit indices (Echeverría et al., 2017; Franken et al., 2018; Lamers et al., 2011; Perugini et al., 2017; Petrillo et al., 2015) it is recommended to do further investigations regarding MHC-SF-R's model fit with a larger sample size and an appropriate estimation method because it might be that the present finding of MHC-SF-R's factorial validity is not reliable. In addition, it should be considered to replace question 6, 8 and 13 belonging to the MHC-SF-R with its original items (see Appendix B) because the original items displayed considerably higher factor loadings.

Since two of the items which displayed low factor loadings (question 6 and 8) belong to MHC-SF-R's societal WB scale, which also displays inacceptable internal consistency ($\alpha = .65$), it seems that MHC-SF-R's new societal WB scale is problematic. Therefore, it is recommended that future research investigates if replacing question 6 and 8 with its unrevised original items yields to better internal consistency and higher factor loadings with regard to MHC-SF-R's societal WB scale. Since the original version of MHC-SF-R's societal WB subscale, that is MHC-SF's SWB subscale (see Appendix B), displayed higher standardized factor loadings and performed notably better in terms of internal consistency

(α = .83), it is expected that exchanging question 6 and 8 leads to higher internal consistency and higher factor loadings in MHC-SF-R's societal WB scale. In case future research reveals that exchanging those questions does not improve factor loadings and internal consistency of MHC-SF-R's societal WB scale, it should be taken into consideration to conduct qualitative research as suggested by Köhle (2010). Köhle (2010) made use of the method "thinking aloud" for finding ambiguous items in the MHC-SF. The same could be done with the MHC-SF-R to explore which problems subjects encounter when filling in MHC-SF-R's societal WB scale, and other subscales belonging to the MHC-SF-R. If further improvements of MHC-SF-R's societal WB scale do not yield to acceptable factor loadings and acceptable internal consistency it can be considered to exchange MHC-SF-R's societal WB scale with MHC-SF's SWB scale. As both scales are based on Keyes's (1998) model of SWB, both scales should measure the same psychological construct and, by extension, exchanging the scales should not have an influence on the measured construct. However, one of the aims of the present study was to improve the original SWB scale by editing it for clarity. Therefore, replacing MHC-SF-R's societal WB scale with MHC-SF's SWB scale should only be taken into consideration if exchanging items and further qualitative research does not improve the scale or reveals useful suggestions how to improve the scale.

With regard to MHC-SF-R's other subscales, it can be said that the new developed relational WB scale displayed high internal consistency (α = .85) and appropriate standardized factor loadings. Therefore, it is suggested to maintain and further investigate the new developed scale. Regarding MHC-SF-R's PWB subscale, factor loadings were generally moderate to high and internal consistency was acceptable. However, the factor loading found for question 13 was low. Since the factor loading of the corresponding unrevised item was higher it should be considered to exchange those items (see Appendix B). Moreover, as internal consistency for MHC-SF's PWB scale was acceptable but not high, it should also be considered to further investigate MHC-SF-R's PWB scale by means of quantitative and qualitative research methods. Concerning MHC-SF-R's EWB subscale, internal consistency and factor loadings were high. Hence, it is recommended to maintain this scale.

In the present study convergent validity between the MHC-SF-R and the RSE and between the MHC-SF-R and the GSE was established. The RSE and GSE are questionnaires investigating aspects of positive functioning in individual life. As a consequence the focus of the present study was on investigating convergent validity between MHC-SF-R's total scale with the validation questionnaires and MHC-SF-R's PWB scale with the validation

questionnaires. However, the current study did not incorporate specific validation questionnaires aiming to examine convergent validity for MHC-SF-R's EWB, societal WB, and relational WB scale. Moreover, the present study did not contain any questionnaire with which discriminant validity between MHC-SF-R's outcome and psychopathology can be determined. Hence, for future research it is recommended to incorporate validation questionnaires for establishing criterion validity that were not addressed in the current study. For examining convergent validity for MHC-SF-R's EWB scale, it is recommended to choose questionnaires measuring life satisfaction or positive affect as suggested by Lamers et al. (2011), e.g. the "Satisfaction with Life Scale" (Pavot & Diener, 1993) or the "Positive Affect Schedule" (Watson, Clark, & Tellegen, 1988). In order to establish convergent validity for MHC-SF-R's societal WB scale, it is suggested to use questionnaires measuring involvement in society. For this purpose Lamers et al. (2011) developed a questionnaire which measures social engagement. Regarding convergent validity for MHC-SF's relational WB subscale it is recommended to use questionnaires measuring trust or affiliation, e.g. the "Propensity to Trust Scale" (Frazier, Johnson, & Fainshmidt, 2013) or the "Need to Belong Scale" (Leary, Kelly, Cottrell, & Schreindorfer, 2013). For determining discriminant validity between MHC-SF-R's outcome of overall mental well-being and other measurements investigating mental illness, it is recommended to use a questionnaire which measures psychopathology, e.g. the "Brief Symptom Inventory" (Derogatis & Melisaratos, 1983).

Moreover, the study design of this study did not allow to investigate MHC-SF-R's test-retest reliability. As a consequence, it was not possible to explore if the MHC-SF-R is sensitive to changes. This is an important issue that should be addressed in further research since before using the MHC-SF-R in intervention research it should be investigated if the MHC-SF-R is sensitive to changes regarding changes in overall mental well-being. In case the MHC-SF-R would not be sensitive to changes, the measurement instrument would fail to measure effects in interventions on mental well-being caused by the independent variable.

4.4 Conclusion

It seems that in the MHC-SF-R not all of MHC-SF's shortcomings are overcome. MHC-SF's SWB items were MHC-SF's most problematic items and the revised SWB items, i.e. MHC-SF-R's societal WB items, are the most problematic items in the MHC-SF-R. Therefore, the revision of the SWB scale seems to be not successful. Regarding the revised PWB scale, it appears that the revision of PWB items did not significantly affect the psychometric properties of the PWB scale. Merely one revised PWB item displayed a striking lower factor loading compared with its unrevised item.

However, the current study also demonstrates that revisions in the MHC-SF-R lead to improvements in the measurement instrument. The finding that MHC-SF-R's EWB scale displayed significantly higher internal consistency than MHC-SF's EWB scale might be an indicator that MHC-SF-R's modified instructions and response format leads to more reliable results than MHC-SF's, since EWB items were the same for both versions of the MHC-SF. Moreover, integrating a relational WB scale into the MHC-SF-R seems to be meaningful because MHC-SF-R's relational WB scale displayed good psychometric properties and internal consistency for MHC-SF-R's overall scale including relational WB items was high. This indicates that the relational WB scale fits into the MHC-SF-R.

Nevertheless, at present, the MHC-SF-R should not be used in intervention research or clinical practice since its societal WB scale lacks internal consistency and present fit indices suggest that MHC-SF-R's model fit is not acceptable. In future research in which shortcomings of the current study regarding the implementation of confirmatory factor analysis are overcome, it remains to be seen if MHC-SF-R's factorial structure is indeed inacceptable or if the MHC-SF-R will show an acceptable model fit.

Since previous research consistently revealed that the MHC-SF shows an acceptable model fit and in the current study internal consistency for all scales belonging to the MHC-SF were found to be at least acceptable, for the time being the MHC-SF should be preferred over the MHC-SF-R. However, if further validation research reveals that MHC-SF-R's factorial structure is acceptable and MHC-SF-R's societal WB scale is further developed so that it consistently displays at least acceptable internal consistency, the MHC-SF-R becomes a more comprehensive measurement instrument for investigating overall mental well-being than the MHC-SF. In this case it might be that the MHC-SF-R supersedes the MHC-SF.

References

Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin, 103*(3), 411-423. doi:10.1037/0033-2909.103.3.411

Baumeister, R. F. (1993). *Self-Esteem: The Puzzle of Low Self-Regard*. New York, USA: Plenum Press.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does High Self-Esteem Cause Better Performance, Interpersonal Success, Happiness, or Healthier Lifestyles? *Psychological Science in the Public Interest, 4*(1), 1-44. doi:10.1111/1529-1006.01431

Bentler, P. M., & Chou, C.-P. (1987). Practical Issues in Structural Modeling. *Sociological Methods & Research, 16*(1), 78-117. doi:10.1177/0049124187016001004

Brown, T. A. (2015). *Confirmatory Factor Analysis for Applied Research* (2nd ed.). New York: The Guilford Press.

Brumfitt, S. M., & Sheeran, P. (1999). The Development and Validation of the Visual Analogue Self-Esteem Scale (VASES). *British Journal of Clinical Psychology, 38*(4), 387-400.

Chou, C.-P., & Bentler, P. M. (1995). Estimates and Tests in Structural Equation Modeling *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 37-55). Thousand Oaks, USA: Sage Publications.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin, 112*(1), 155-159. doi:10.1037/0033-2909.112.1.155

Cronbach, L. J. (1951). Coefficient Alpha and the Internal Structure of Tests. *Psychometrika, 16*, 297-334. doi:10.1007/BF02310555

Derogatis, L. R., & Melisaratos, N. (1983). The Brief Symptom Inventory: An Introductory Report. *Psychological Medicine, 13*(3), 595-605.

Diedenhofen, B., & Musch, J. (2015). Cocor: A Comprehensive Solution for the Statistical Comparison of Correlations. *PLOS ONE, 10*(4), e0121945. doi:10.1371/journal.pone.0121945

Diedenhofen, B., & Musch, J. (2016). Cocron: A Web Interface and R Package for the Statistical Comparison of Cronbach's Alpha Coefficients. *International Journal of Internet Science, 11*(51-60).

Diener, E., Suh, E. M., Lucas, R., & Smith, H. L. (1999). Subjective Well-Being: Three Decades of Progress. *Psychological Bulletin, 125*(2), 276-302.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of Estimation Methods, Number of Indicators per Factor, and Improper Solutions on Structural Equation Modeling Fit Indices. *Structural Equation Modeling: A Multidisciplinary Journal, 2*(2), 119-143. doi:10.1080/10705519509540000

Echeverría, G., Torres, M., Pedrals, N., Padilla, O., Rigotti, A., & Bitran, M. (2017). Validation of a Spanish Version of the Mental Health Continuum-Short Form Questionnaire. *Psicothema, 29*(1), 96-102. doi:10.7334/psicothema2016.3

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement, 11*(1), 93-103. doi:10.1177/014662168701100107

Fisher, R. A. (1925). *Statistical Methods for Research Workers* (11th ed.). Edinburgh: Oliver and Boyd.

Franken, K., Lamers, S. M. A., Ten Klooster, P. M., Bohlmeijer, E. T., & Westerhof, G. J. (2018). Validation of the Mental Health Continuum-Short Form and the Dual Continua Model of Well-Being and Psychopathology in an Adult Mental Health Setting. *Journal of Clinical Psychology, 74*(12), 2187-2202. doi:10.1002/jclp.22659

Frazier, M. L., Johnson, P. D., & Fainshmidt, S. (2013). Development and Validation of a Propensity to Trust Scale. *Journal of Trust Research, 3*(2), 76-97. doi:10.1080/21515581.2013.820026

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational Research: An Introduction*: Pearson/Allyn & Bacon.

Gecas, V. (1989). The Social Psychology of Self-Efficacy. *Annual Review of Sociology, 15*, 291-316.

Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An Item Response Theory Analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 23*(5), 443-451. doi:10.1177/0146167297235001

Greenspoon, P. J., & Saklofske, D. H. (2001). Toward an Integration of Subjective Well-Being and Psychopathology. *Social Indicators Research, 54*(1), 81-108. doi:10.1023/A:1007219227883

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson University Press.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association, 82*(400), 1147-1149. doi:10.1080/01621459.1987.10478551

Hu, L., & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Kahneman, D. (1973). *Attention and Effort*: Prentice-Hall.

Keyes, C. L. M. (1998). Social Well-Being. *Social Psychology Quarterly, 61*(2), 121-140. doi:10.2307/2787065

Keyes, C. L. M. (2002). The Mental Health Continuum: From Languishing to Flourishing in Life. *Journal of Health and Social Research, 43*(2), 207-222.

Keyes, C. L. M. (2005a). Mental Illness and/or Mental Health? Investigating Axioms of the Complete State Model of Health. *Journal of Consulting and Clinical Psychology, 73*(3), 539-548. doi:10.1037/0022-006X.73.3.539

Keyes, C. L. M. (2005b). The Subjective Well-Being of America's Youth: Toward a Comprehensive Assessment. *Adolescent & Family Health, 4*(1), 3-11.

Keyes, C. L. M., 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 and Psychotherapy, 15*(3), 181-192. doi:10.1002/cpp.572

Kim, H. Y. (2013). Statistical Notes for Clinical Researchers: Assessing Normal Distribution (2) Using Skewness and Kurtosis. *Restorative Dentistry & Endodontics, 38*(1), 52-54. doi:10.5395/rde.2013.38.1.52

Kline, P. (2000). *Handbook of Psychological Testing*: London: Routledge.

Köhle, N. (2010). *Mag ik vragen, wat ik u moet vragen?* (Bachelor's Degree), University of Twente, Enschede.

Lamers, S. M. A., Westerhof, G. J., Bohlmeijer, E. T., Ten Klooster, P. M., & Keyes, C. L. M. (2011). Evaluating the Psychometric Properties of the Mental Health Continuum-Short Form (MHC-SF). *Journal of Clinical Psychology, 67*(1), 99-110. doi:10.1002/jclp.20741

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct Validity of the Need to Belong Scale: Mapping the Nomological Network. *Journal of Personality Assessment, 95*(6), 610-624. doi:10.1080/00223891.2013.819511

Li, C. H. (2016). Confirmatory Factor Analysis with Ordinal Data: Comparing Robust Maximum Likelihood and Diagonally Weighted Least Squares. *Behavior Research Methods, 48*(3), 936-949. doi:10.3758/s13428-015-0619-7

Luszczynska, A., Scholz, U., & Schwarzer, R. (2005). The General Self-Efficacy Scale: Multicultural Validation Studies. *The Journal of Psychology, 139*(5), 439-457. doi:10.3200/JRLP.139.5.439-457

McCurdy, B. A., & Kelly, D. B. (1997). Correlations of the MMPI-2 Low Self-Esteem Scale with Two Self-Esteem Measures. *Psychological Reports, 81*(3), 826-826. doi:10.2466/pr0.1997.81.3.826

Mortel, T. F. (2008). Faking It: Social Desirability Response Bias in Self-Report Research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

Murre, J. M., & Dros, J. (2015). Replication and Analysis of Ebbinghaus' Forgetting Curve. *PLOS ONE, 10*(7), e0120644. doi:10.1371/journal.pone.0120644

Orpana, H., Vachon, J., Dykxhoorn, J., & Jayaraman, G. (2017). Measuring Positive Mental Health in Canada: Construct Validation of the Mental Health Continuum-Short Form. *Health Promotion and Chronic Disease Prevention in Canada, 37*(4), 123-130. doi:10.24095/hpcdp.37.4.03

Pavot, W., & Diener, E. (1993). Review of the Satisfaction With Life Scale. *Psychological Assessment, 5*(2), 164-172. doi:10.1037/1040-3590.5.2.164

Perugini, M. L. L., Iglesia, G., Castro, A. S., & Keyes, C. L. M. (2017). The Mental Health Continuum–Short form (MHC–SF) in the Argentinean Context: Confirmatory Factor Analysis and Measurement Invariance. *Europe's Journal of Psychology, 13*(1), 93-108. doi:10.5964/ejop.v13i1.1163

Petrillo, G., Capone, V., Caso, D., & Keyes, C. L. M. (2015). The Mental Health Continuum-Short Form (MHC–SF) as a Measure of Well-Being in the Italian Context. *Social Indicators Research, 121*(1), 291-312. doi:10.1007/s11205-014-0629-3

Prilleltensky, I. (2005). Promoting Well-Being: Time for a Paradigm Shift in Health and Human Services. *Scandinavian Journal of Public Health, 33*(66), 53-60. doi:10.1080/14034950510033381

Robins, R. W., Hendin, H. M., & Trzesniewski, K. H. (2001). Measuring Global Self-Esteem: Construct Validation of a Single-Item Measure and the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin, 27*(2), 151-161. doi:10.1177/0146167201272002

Rosenberg, M. (1979). *Conceiving the Self*: New York: Basic Books.

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. doi:10.1146/annurev.psych.52.1.141

Ryff, C. D. (1989). Happiness Is Everything, or Is It? Explorations on the Meaning of Psychological Wellbeing. *Journal of Personality and Social Psychology, 57*(6), 1069-1081. doi:10.1037/0022-3514.57.6.1069

Satorra, A., & Bentler, P. M. (1994). Corrections to Test Statistics and Standard Errors in Covariance Structure Analysis *Latent Variables Analysis: Applications for Developmental Research* (pp. 399-419). Thousand Oaks, USA: Sage Publications.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research, 8*(2), 23-74.

Schmitt, D. P., & Allik, J. (2005). Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem. *Journal of Personality and Social Psychology, 89*(4), 623-642. doi:10.1037/0022-3514.89.4.623

Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is General Self-Efficacy a Universal Construct? Psychometric Findings From 25 Countries. *European Journal of Psychological Assessment, 18*(3), 242-251. doi:10.1027//1015-5759.18.3.242

Schwarzer, R. (1992). *Self-Efficacy: Thought Control of Action*. Washington DC: Hemisphere.

Schwarzer, R., Bäßler, J., Kwiatek, P., Schröder, K., & Zhang, J. X. (1997). The Assessment of Optimistic Self-Beliefs: Comparison of the German, Spanish, and Chinese Versions of the General Self-Efficacy Scale. *Applied Psychology: An International Review, 46*(1), 69-88. doi:10.1111/j.1464-0597.1997.tb01096.x

Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy Scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in Health Psychology: A User's Portfolio. Causal and Control Beliefs* (pp. 35-37). Windsor, England: NFER-Nelson.

Schwarzer, R., & Jerusalem, M. (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen [Scales for the assessment of teacher and student characteristics]*. Berlin, Germany: Freie Universität Berlin.

Torrey, W. C., Mueser, K. T., McHugo, G. H., & Drake, R. E. (2000). Self-Esteem as an Outcome Measure in Studies of Vocational Rehabilitation for Adults with Severe Mental Illness. *Psychiatric Services, 51*(2), 229-233. doi:10.1176/appi.ps.51.2.229

Vuletić, G., Erdeši, J., & Nikić, L. B. (2018). Factor Structure and Validation of Croatian Version Questionnaire of Mental Health Continuum MHC-SF. *Medica Jadertina, 48*(3), 113-124.

Wang, X., Zhang, D., & Wang, J. (2011). Dual-Factor Model of Mental Health: Surpass the Traditional Mental Health Model. *Psychology, 2*(8), 767-772. doi:10.4236/psych.2011.28117

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. doi:10.1037/0022-3514.54.6.1063

Westerhof, G. J. (2019) *MHC-SF: Reoccuring Difficulties in Item's Interpretation/Interviewer: A. Dostall*.

Westerhof, G. J., & Keyes, C. L. M. (2010). Mental Illness and Mental Health: The Two Continua Model Across the Lifespan. *Journal of Adult Development, 17*(2), 110-119. doi:10.1007/s10804-009-9082-y

Appendix A

MHC-SF (ages 18 years or older)

Theoretical Dimension	Number in MHC-SF	In the past month, how often did you feel:
EWB [based on (Diener et al., 1999)]		
Happiness	1	Happy
Interest	2	Interested in life
Life Satisfaction	3	Satisfied with life
PWB [based on (Ryff, 1989)]		
Self-Acceptance	9	That you liked most parts of your personality
Mastery	10	Good at managing the responsibilities of your daily life
Positive relations	11	That you had warm and trusting relationships with others
Personal growth	12	That you have experiences that challenge you to grow and become a better person
Autonomy	13	Confident to think or express your own ideas and opinions
Purpose in life	14	That your life has a sense of direction or meaning to it
SWB [based on (Keyes, 1998)]		
Soc. contribution	4	That you had something important to contribute to society
Soc. integration	5	That you belonged to a community (like a social group, your neighborhood, your city)
Soc. actualization	6	That our society is becoming a better place for people
Soc. acceptance	7	That people are basically good
Soc. coherence	8	That the way our society works makes sense to you

Note. In the MHC-SF subjects have to indicate on a 6-point-Likert-Scale ranging from 0 to 5 how they have been feeling during the past month in which 0 ≙ never, 1 ≙ once or twice, 2 ≙ about once a week, 3 ≙ about 2 or 3 times a week, 4 ≙ almost every day, 5 ≙ every day. MHC-SF = Mental Health Continuum Short-Form; EWB = emotional well-being; PWB = psychological well-being; SWB = social well-being; soc. = social.

Appendix B

MHC-SF-R (ages 18 years or older)

Theoretical Dimension	Number in MHC-SF-R	Corresponding item in MHC-SF	In the past week, how often did you feel:
EWB [based on (Diener et al., 1999)]			
Happiness	1	1	I am happy
Interest	2	2	I am interested in life
Life Satisfaction	3	3	I am satisfied with life
PWB [based on (Ryff, 1989)]			
Self-Acceptance	9	9	I accept myself as I am.
Mastery	10	10	I am able to master my life.
Positive relations	11	11	I share love and sorrow with some people.
Personal growth	12	12	I can develop myself.
Autonomy	13	13	I stand up for myself.
Purpose in life	14	14	I feel my life has purpose.
Societal WB [based on (Keyes, 1998)]			
Soc. contribution	4	4	I make a valuable contr. to our society
Soc. integration	7	5	I belong to a group of people.
Soc. integration	19	-	I find my place in this society.
Soc. actualization	5	6	I think our country is developing well.
Soc. acceptance	6	7	I accept others as they are.
Soc. coherence	8	8	I understand how our society works.
Relational WB [based on (Prilleltensky, 2005)]			
Rel. affirmation	15	-	I can mean something for others.
Rel. affection	17	-	I feel connected to other people.
Rel. satisfaction	16	-	I am satis. with my social contacts.
Rel. trust	18	-	I can rely on other people.

Note. In the MHC-SF-R subjects have to indicate on a 6-point-Likert-Scale ranging from 0 to 5 how they have been feeling during the past week in which 0 ≙ Never, 1 ≙ rarely, 2 ≙ sometimes, 3 ≙ regularly, 4 ≙ often, 5 ≙ almost always. MHC-SF-R = Mental Health Continuum Short Form revised; EWB = emotional well-being; PWB = psychological well-being; WB = well-being; soc. = social; rel. = relational; satis. = satisfied; contr. = contribution.

Appendix C

Informed consent

Informed consent

Your participation in this study is entirely voluntary and you can withdraw at any time, without having to state a reason.

We believe there are no known risks associated with this research study; however, as with any online related activity the risk of a breach is always possible. Your responses to the survey will be kept confidentially and anonymously. We will minimize any risks by storing your data safely, without offering access to third parties.

Study contact details for further information:

m.boggemann@student.utwente.nl
a.dostall@student.utwente.nl
b.nijhuis@student.utwente.nl
l.wenglorz@student.utwente.nl

Contact Information for Questions about Your Rights as a Research Participant

If you have questions about your rights as a research participant, or wish to obtain information, ask questions, or discuss any concerns about this study with someone other than the researcher(s), please contact the Secretary of the Ethics Committee of the Faculty of Behavioral, Management and Social Sciences at the University of Twente
by: ethicscommittee-bms@utwente.nl

I have read the above mentioned terms and hereby agree to voluntarily participate in this study.

Appendix D

Testing data for normality

The normality test Shapiro-Wilk indicated for MHC-SF-R's total scale (p = .027), its EWB scale (p = .006), its societal WB scale (p = .018) and its relational WB scale (p = .010) that data is not normally distributed. For MHC-SF's total scale (p = .007), its EWB scale (p = .003) and its SWB scale (p = .023) the Shapiro-Wilk test also illustrated that data is not normally distributed. For GSE's total scale (p = .172), RSE's total scale (p = .124), MHC-SF-R's PWB scale (p = .130) and MHC-SF's PWB scale (p = .051), the Shapiro-Wilk test suggested that data is normally distributed. However, the Shapiro-Wilk test is quite strict in checking for normality. For this reason, a second test for examining normality using skewness and kurtosis suggested by Kim (2013) was implemented, namely, a Z-test for which Z-scores are obtained by dividing skew values and excess kurtosis values by their standard errors. According to Kim (2013), for medium samples ($50 < N < 300$) the null hypothesis is rejected at Z-values over 3.29 which complies with $\alpha = .05$. According to this method, all scales in the current study are normally distributed. However, it is important to take into consideration that both groups are tightly above 50 subjects. Therefore, setting the cutoff at a Z-value over 3.29 is quite generous. Due to the contradictory normality test results, additionally histograms with normality curves and Q-Q plots were examined. The eyeball test confirmed that data approaches normally distribution and Q-Q plots revealed that some minor outliers (interquartile range multiplied with the factor 1.5) are present which might be the reason why the Shapiro-Wilk test was significant for some scales. However, due to the fact that those outliers were not identified by means of the outlier labeling rule defined by Hoaglin and Iglewicz (1987), they were maintained for analyses. The results of the various normality tests indicate that data for all subscales is at least approximately normally distributed.