Psychometric Evaluation of a Revised Practice-Friendly Version of the Mental Health Continuum Short Form

Julia Wulf

s2952084

University of Twente

BMS Faculty

Department of Psychology

Course 202001489: Master Thesis

Supervisor: Dr. Erik Taal

Second Supervisor: Prof. Dr. Gerben Westerhof

5th July 2023

Abstract

Background: This exploratory paper aimed at investigating the psychometric properties of two revised versions of the Mental Health Continuum-Short Form (MHC-SF), measuring mental well-being, in terms of their factor structure, internal consistency and distribution of item and scale scores. Both revised versions include reformulated items and four new items assessing social relationships while one (RQ-OF) retained the original response format (e.g., monthly, specific estimation of feelings) and the other (RQ-RF) includes an adapted response format (e.g., weekly, rough estimation of feelings). **Methods:** The used data was obtained from the LISS-panel, representing a true probability sample from the Netherlands. Analyses of the factor analysis, Cronbach's alpha, ceiling, and floor effects were conducted. **Results:** Both the RQ-OF and RQ-RF had a similar but poor fit on the three-factor structure. The internal consistency of both versions was excellent. On an item level, the RQ-RF displayed fewer ceiling effects and was hence more normally distributed than the RQ-OF. On a scale level, both were similarly normally distributed. **Discussion:** The change in response format only improved item distributions, with no improvements in factor structure, internal consistency, and scale distribution. Therefore, it remains inconclusive to determine the superiority of either version in terms of psychometric properties. Additional investigations are needed to validate the observed findings and to gain a more comprehensive understanding of the impact of response format changes on the psychometric properties.

Psychometric Evaluation of a Revised Practice-Friendly Version of the Mental Health Continuum Short Form

In recent years, there has been a notable change in the field of psychology as mental well-being has been incorporated into the definition of mental health (Iasiello et al., 2020). Mental well-being refers to a positive state of psychological functioning, characterized by subjective feelings of happiness, life satisfaction, and fulfilment (Keyes, 2002). The expansion of the definition of mental health has led to treatments aimed at improving positive functioning, with well-being interventions showing promising results in increasing mental health (Lamers et al., 2011). Studies have demonstrated that such interventions can increase well-being, improve psychological distress, hasten recovery from severe mental illness and even prevent the onset of mental illness altogether (Chakhssi et al., 2018; Fava et al., 2017; Jeste et al., 2015). To effectively assess mental well-being, a reliable and valid tool is essential for guiding intervention and conducting further research. Indeed, Iasiello et al. (2020) emphasised the importance of developing and improving a mental well-being scale in clinical practice, such as the Mental Health Continuum-Short Form (MHC-SF), to achieve these goals. However, feedback from participants has suggested that the MHC-SF questionnaire may not be user-friendly in various parts. This study aims to evaluate and compare two revised versions of the MHC-SF questionnaire to improve its usability. Before doing so, the upcoming sections will introduce the concept of mental health, including mental well-being and mental illness, to foster a deeper comprehension of the topic.

Conceptualizing Mental Health

Historically, the medical model has dictated the understanding of mental health, focusing on the absence of disease or illness rather than promoting well-being (Huda, 2021). However, this view has changed in recent decades with the advancement of research in positive psychology and the study of mental well-being. Thus, the initial definition was challenged and linked to mental well-being (Keyes, 2002). The World Health Organization

revised the definition of mental health to include "a state of well-being in which the individual realises his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" (World Health Organization, 2004, p. 12). This shift paved the way to include positive interventions in clinical practice aimed at increasing mental well-being *in addition* to alleviating distress (Jeste et al., 2015; Leamy et al., 2011). The significance of mental well-being being included within the definition of mental health and the need for developing interventions that improve positive functioning has become increasingly recognized in recent years (Seligman & Csikszentmihalyi, 2000). To explore this further, two conceptual models, namely the two continua model and the tripartite model, are introduced to provide a deeper understanding of mental health and mental well-being.

Two Continua Model

The two continua model, developed by Keyes (2005) and further elaborated by Westerhof and Keyes (2010), proposes that mental well-being and mental illness are distinct but interrelated concepts. In particular, the model suggests that the absence of mental illness does not automatically indicate the presence of mental health. Instead, mental health is best understood as the existence of mental well-being and the absence of mental illness. Thus, to comprehensively assess mental health according to their model, it is necessary to measure both mental well-being and mental illness (Lamers et al., 2011). Mental well-being can be further categorised into three dimensions, known as the tripartite model.

The Tripartite Model

Nearly half a century of research has yielded the tripartite model of well-being consisting of three dimensions, namely emotional (EWB), psychological (PWB), and social well-being (SWB; Joshanloo, 2022; Lamers et al., 2011). These dimensions can either be categorized as hedonic or eudaimonic well-being (Keyes, 2006; Keyes & Waterman, 2003).

Hedonic well-being includes feelings of pleasure and happiness (Waterman, 1993) and can be enhanced by creating a high balance of pleasant to unpleasant feelings (Lamers et al., 2011). This is reflected in the EWB dimension, measuring the presence of positive affect and life satisfaction (Keyes, 2002; Lamers et al., 2011). In contrast, the eudaimonic orientation goes beyond happiness as it encompasses self-realization and personal expression, including advancing one's skills and goals (Waterman, 1993). This orientation is associated with the two dimensions of well-being, PWB and SWB (Keyes, 2002). PWB refers to an individual's sense of thriving and includes dimensions such as self-acceptance, positive relations, autonomy, environmental mastery, purpose in life, and personal growth (Keyes, 2002; Ryff, 1989). On the other hand, SWB describes an individual's perception of their social environment, such as social contribution, integration, actualization, acceptance, and coherence (Keyes, 1998, 2002). Therefore, the eudaimonic orientation and its associated dimensions provide a more comprehensive understanding of mental well-being beyond just the presence of happiness. To comprehensively assess mental well-being, researchers have developed various scales and measurement tools that aim to capture the different dimensions of mental well-being.

Assessing Mental Well-Being

To effectively measure an individual’s state of mental well-being, as described in the two continua model, a scale is needed that includes exclusively the tripartite model with its three dimensions EWB, SWB, and PWB. So far, various scales and questionnaires exist to assess the tripartite model, including the Mental Health Continuum-Long Form (Keyes & Magyar-Moe, 2003). This questionnaire has been validated and proven reliable in measuring all three dimensions of mental well-being, but the 40-item scale requires a significant amount of time and energy to complete (Lamers et al., 2011). Since other questionnaires also exhibit certain issues such as only assessing one dimension of the model or including items on mental illness, the Mental Health Continuum-Short Form (MHC-SF) was developed to address these issues.

MHC-SF: Characteristics and Limitations

The development of the MHC-SF aimed to create a more practical but still valid and reliable questionnaire that covers all three dimensions of mental well-being (Keyes et al., 2008; Lamers et al., 2011), even across translations and cultures (Joshanloo et al., 2013). The MHC-SF consists of 14 items, with each item representing one aspect of the three dimensions of mental well-being. Specifically, three items assess EWB, six items assess PWB, and five items assess SWB. In an evaluative study by Lamers et al. (2011), the MHC-SF demonstrated acceptable internal consistency ($\alpha > 0.80$), test-retest reliability, and convergent validity. Due to its enhanced user-friendliness and shorter administration, it has become the most commonly used instrument of mental well-being (Iasiello et al., 2022).

Nevertheless, the MHC-SF also has its shortcomings. Firstly, the factor structure of the MHC-SF has been a matter of debate, with different studies suggesting various factor solutions. Iasiello et al. (2022) conducted a meta-analysis and identified five different factor structures for the MHC-SF, namely one-factor, two-factor, three-factor, hierarchical, and bifactor models. A one-factor model suggests that all the variances would be explained best by a single mental well-being factor. A two-factor structure may represent mental well-being as the predominant factor and an additional underlying concept as the second factor, together explaining all the variance. A three-factor structure is often chosen, as it is mostly in line with the current research of theory, as the three dimensions of well-being each represent one factor. A hierarchical model includes an additional second-order mental well-being factor, while a bifactor model integrates both the one-factor and three-factor models, assuming both models independently influence the items (Iasiello et al., 2022).

A study conducted by Lamers et al. (2011) examined the various models for the MHC-SF and determined that the three-factor model, consisting of EWB, SWB, and PWB, provided the best fit. In their confirmatory factor analysis, these three factors accounted for 58% of the variance. This finding was further supported by a meta-analysis conducted by Iasiello et al.

(2022), which examined factor structures in 26 studies. The meta-analysis reinforced the conclusion that the three-factor model was not only the best fit for the MHC-SF but also aligned with the existing literature on the topic. Since a scale with a poor factor structure may potentially lead to inaccurate assessments of mental health, this study will review the factor structure in the analysis.

An additional limitation of the MHC-SF highlighted in previous research is that the SWB subscale of the MHC-SF has been found to have relatively lower reliability compared to EWB and PWB subscales (Köhle, 2010). This may affect the accuracy of the scale in assessing mental well-being and will be addressed in the present study.

The response format of the MHC-SF has also been reported to have some limitations. Specifically, participants have reported to have found it difficult to estimate the specific frequencies of certain feelings during the past month (e.g., "2-3 times a week"; Geisen, 2020; Köhle, 2010). Additionally, participants found that some items were difficult to answer as they were formulated too abstractly (e.g., "how often do you feel that society is becoming a better place for you?"; Köhle, 2010). Lastly, participants expressed that the SWB scale lacks items on closer social relations and interpersonal well-being, as so far, it only includes items on perceived societal functioning (Köhle, 2010).

MHC-SF: Revisions

To address the reported limitations of the MHC-SF and make the questionnaire more user-friendly and comprehensive, certain elements of the questionnaire were revised by a research team from the University of Twente, the Netherlands, to improve the usability of the questionnaire. Firstly, the response format was changed from the past month to the past week. Additionally, instead of asking for specific details on the frequencies of feelings, rough estimates were used, namely never, seldom, sometimes, regularly, often, and (almost) always. These changes were made to reduce a possible recall bias, referring to the systematic errors that can occur when collecting data from study participants because they cannot accurately

recall past events or experiences. It can affect the validity and reliability of research findings because it can lead to under- or overestimation of certain factors (Gotlin et al., 2020). Hence, by allowing participants to select a rather rough response option and shortening the reference period, a greater response accuracy can be achieved (Martin, 2005).

Secondly, items were reformulated to be simpler and intelligible to all. The adapted items were approved by the creator of the instrument, prof. Corey Keyes. Lastly, four items were added to the SWB scale assessing social well-being in social relations including contribution, satisfaction, relatedness, and support.

The changes in response format in the MHC-SF questionnaire may have affected the presence of floor and ceiling effects. These effects describe how frequently participants select the lowest or highest score on the scale. If over 15% of participants choose either end of the scale, it indicates a floor or ceiling effect. These effects prevent that participants' answers can be differentiated at the extremes of the scales. As a result, the reliability may be compromised (Terwee et al., 2007). This study will be one of the first to assess potential floor and ceiling effects and will provide new insights into the structure of the MHC-SF.

The Present Study

This study aims at investigating and comparing the psychometric properties of two versions of the revised MHC-SF. The first version consists of the revised questionnaire with the original response format (RQ-OF). It includes the changes of the item reformulation and added SWB items while retaining the original response format (e.g., past month, specific estimation of feelings). The second version consists of the revised questionnaire with revised response format (RQ-RF). It includes the changes of item reformulation, added social items and the revised response format (e.g., past week, rough estimation of feelings).

The aim of the present study is to examine the effects of the revisions made to the mental well-being measure on its psychometric properties, with the ultimate goal of

enhancing its user-friendliness. To achieve this aim, the study will examine the following research questions:

RQ1: What is the factor structure of the RQ-RF in comparison to the RQ-OF?

RQ2: What is the internal consistency of the RQ-RF in comparison to the RQ-OF?

RQ3: What is the distribution of the scores on an item-level and scale-level and how do they and the mean scores differ between RQ-ROF and RQ-RF?

Methods

Design and Procedure

The present study made use of an exploratory and non-experimental design. The used data was obtained from the Longitudinal Internet Studies for the Social Sciences (LISS) panel, provided by the independent non-profit research institute Centerdata from Tilburg University in the Netherlands. The LISS panel represents a true probability sample of households in the Netherlands, hence ensuring its representativeness and is commonly used for scientific, social, and policy-relevant research. In total, the LISS panel consists of 5000 households, equaling approximately 7500 individuals, who complete online questionnaires every month in exchange for financial compensation for each completed questionnaire. The questionnaires cover eight core themes, namely health, religion and ethnicity, social integration and leisure, family and household, work and schooling, personality, politics and values, economic situation: assets, income, and housing (LISS panel, 2022).

This study makes use of the health theme of the questionnaires. The data was collected as part of an ongoing well-being study by Westerhof and ten Klooster from the faculty of Behavioral, Management and Social Sciences of the University of Twente. For that, 3571 participants from the LISS panel were randomly divided into four equal groups with similar age and gender distributions. Each group was randomly assigned and requested to complete one of the four versions of the MHC-SF and MHC-SF-R, namely (1) original questionnaire with original response format, (2) original questionnaire with revised response format, (3)

revised questionnaire with original response format (RQ-OF) and (4) revised questionnaire with revised response format (RQ-RF). For the present study, the data of the two latter versions, RQ-OF and RQ-RF, were used. The data was collected in May 2020.

Participants

Participation in the study on mental well-being was voluntary. Out of the initial 3571 participants from the LISS panel, a total of 2714 participants successfully submitted their assigned questionnaires. Five participants did not complete the questionnaire and were consequently excluded from the analysis. Among the remaining participants, 1377 completed versions (1) and (2), while version (3) RQ-OF was completed by 680 participants, and version (4) RQ-RF was completed by 662 participants. For this study, data from version (3) RQ-OF and (4) RQ-RF were utilized, resulting in a total sample size of 1342 participants for the analyses. Since the surveys were conducted in Dutch, all participants had sufficient language levels in Dutch. An overview of the demographics such as age, gender, and highest education of the two versions can be found in Table 1. The gender of the participants was approximately equally distributed with half being female (51.8% in RQ-OF, 52.3% in RQ-RF) and half being male (48.2% in RQ-OF, 47.7 % in RQ-RF) with an age range from 18 to 93 and a mean of 48 years. The highest education for most participants was higher vocational education for both RQ-OF (26.7%) and RQ-OF (23.4%). Further, most participants were either married (43.8% in RQ-OF, 44.1% in RQ-RF) or never had been married (40.4% in RQ-OF, 40.5% in RQ-RF). Moreover, most participants reported being employed (45.1% in RQ-OF, 44.1% in RQ-RF). Overall, it is noteworthy that both versions are highly comparable in all demographic characteristics as the corresponding inferential tests showed no significant differences at an alpha level of .05.

Table 1

Demographics of the RQ-OF (N= 680) and RQ-RF (N= 662) Versions

Characteristic	RQ-OF			RQ-RF			$X^2(df)$	*p*
	M(*SD*)	*n*	%[a]	*M*(*SD*)	*n*	%[a]		
Age (years)	48.0(18.7)			48.7(19.7)			0.67(1332)[b]	.50
Gender							0.34(1)	.85
Male		328	48.2		316	47.7		
Female		352	51.8		346	52.3		
Highest Education							1.82(5)	.87
Primary school		36	5.3		40	6.1		
vmbo		130	19.2		119	18		
havo/vwo		85	12.6		87	13.2		
mbo		162	23.9		155	23.2		
hbo		181	26.7		167	23.4		
wo		83	12.3		93	25.3		
Marital Status							0.34(4)	.99
Married		298	43.8		292	44.1		
Separated		2	0.3		2	0.3		
Divorced		69	10.1		62	9.4		
Widowed		36	5.3		38	5.7		
Never married		275	40.4		268	40.5		
Work Status							3.45(5)	.63
Employed		307	45.1		292	44.1		
Self-employed		41	6.0		31	4.7		
Student		82	12.1		72	10.9		
Retired		133	19.6		147	22.2		
Unemployed		67	9.9		73	9.5		
Other		50	7.4		57	8.6		

Note. Education levels in the Netherlands: vmbo: intermediate secondary education, havo/vwo: higher secondary education, mbo: intermediate vocational education, hbo: higher vocational education, wo: university.

[a]Percentage within the group.

[b]Depicts a *t*-test (*df*).

Instruments

Demographics

Participants' demographic information was assessed using self-constructed items from the LISS panel. Within the LISS panel, the participants answered 27 items on their gender, household, age, year of birth, marital status, housing, place of living, work, income, and education. For this analysis, only the data on gender, age, education, marital status, and work status were used. This served the purpose of determining whether the two samples were comparable. The participants were presented with different answer options and were asked to choose the most suitable one.

MHC-SF Revised

The revised MHC-SF encompasses a total of 18 items. A total of 14 items of the MHC-SF were reformulated and four new items were added to the SWB subscale. The two versions, RQ-OF and RQ-RF, differ in the phrasing of the introduction question (RQ-OF: "In the past month, how often did you have the following feelings?"; RQ-RF: "In the past week, how often did you have the following feelings?"). The EWB subscale contains three items (e.g., "I am interested in life."), the SWB subscale contains nine items (e.g., "I belong to a group of people.") and the PWB subscale encompasses six items (e.g., "I can develop myself.").

Two different response formats were used. The RQ-OF used (0) never, (1) once or twice a month, (2) about once a week, (3) 2 or 3 times a week, (4) almost every day, (5) every day. The RQ-RF used namely (0) never, (1) rarely, (2) sometimes, (3) regularly, (4) often, (5) (almost) always. The scale scores are calculated by averaging the item scores. Details of the items and response options are provided in Appendix A.

Data Analysis

The analysis was conducted with the statistical software platform SPSS (IBM Corp, 2021). After screening the data, five participants who did not complete the questionnaire,

were excluded from both descriptive and inferential analysis. Next, the data was averaged based on the MHC-SF manual and new variables were created by dividing the items into the three different scales of EWB, SWB and PWB based on the manual of the MHC-SF (Keyes, 2002).

For the descriptive analysis, the mean, standard deviation, and the item frequency were used to determine the distribution of the demographics of each group. Additionally, independent t-tests were performed to determine if there were any significant differences between the RQ-OF and RQ-RF versions. To determine the comparability of the two participant groups, chi-square tests were performed to evaluate the comparability in terms of their gender, highest education, marital status, and work status.

For the first research question, an exploratory factor analysis was performed for each version, RQ-OF and RQ-RF. Specifically, a principal component analysis was conducted with a direct oblimin rotation. Eigenvalues greater than one (Kaiser, 1960) and scree-tests (Cattell, 1966) were used to decide on the number of components to extract for each version. Additionally, the two versions were visually compared for the number of extracted components and then for explained variance by the components and factor loadings to examine the fit of the factor structure.

For the second research question, the internal consistency was examined by computing Cronbach's alpha and the item-total correlations at both the item and subscale levels for both RQ-OF and RQ-RF. The differences in the Cronbach's alpha values were determined by examining whether the 95% confidence intervals overlapped. If there was no overlap in the confidence intervals, the alpha values were considered significantly different.

To answer the third research question, the distribution of the item scores for both versions was analysed first by calculating and comparing their skewness and kurtosis to determine the response distribution and normality. They were considered normal if the statistics were between +1 and -1. Next, the item frequencies were examined for ceiling and

floor effects. These were assumed to be present if 15% or more of the participants gave the lowest or highest response (Terwee et al., 2007). The two versions were compared in terms of their floor and ceiling effects. After that, in case of a ceiling or floor effect, chi-square tests were performed to test if the frequencies differed between the two versions. Thereafter, the distribution of the scale scores were analysed using the Skewness and Kurtosis measure and subsequently, the scale scores were analysed for ceiling and floor effects. They were assumed to be present if the values of the scale scores were either 5 or 1. Next, the mean scores and standard deviations for the scales were calculated and t-tests used to test the differences in mean scores between the two versions.

Results

RQ1: Factor Structure

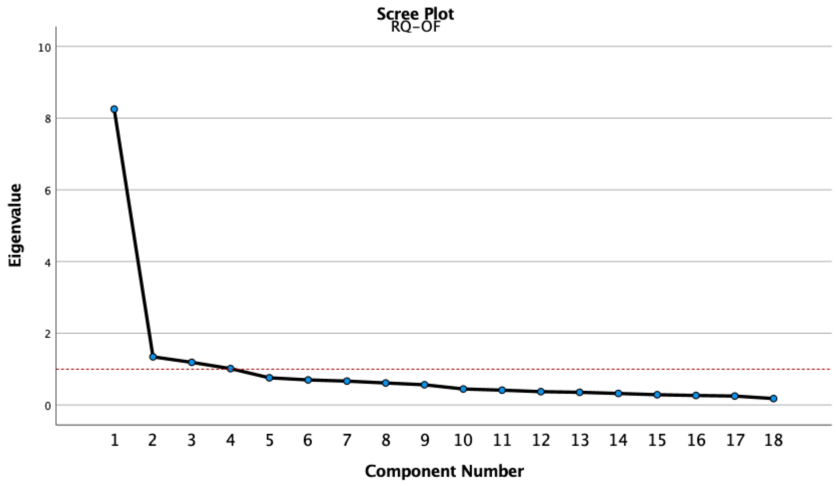
To explore the factor structure of both versions, an exploratory factor analysis with a direct oblimin rotation and scree-test was conducted. The scree-plots were analysed using the elbow method, indicating a sharp drop between components one and two for both versions. This suggests that a one-factor structure may be the most suitable for both questionnaire versions. The scree-plots for RQ-OF and RQ-RF can be found in Figure 1 and 2, respectively.

However, the Eigenvalue results suggested a four-component factor structure for RQ-OF and a three-component factor structure for RQ-RF. Specifically, for the RQ-OF, most items demonstrated higher loadings on the first factor, which accounted for 45.8% of the total variance Adding the second factor increased the total explained variances to 53.3%, while the third factor increased it to 59.9%. Including the fourth factor explained 65.6% of the total explained variance. In contrast, for the RQ-RF, the analysis of individual factor loadings revealed that most items loaded onto the first factor, which accounted for nearly half (47.9%) of the total variance. The inclusion of the second factor increased the total explained variance to 55.4%. Furthermore, the addition of the third factor contributed to a total explained

variance of 61.3%. Given the discrepancy in factor structure suggestions, with the elbow method indicating a one-factor structure and Eigenvalues suggesting a three or four-factor structure, a decision had to be made regarding the appropriate factor structure to use. To align with the existing literature supporting the theoretical background of the three dimensions of mental well-being (EWB, SWB, PWB) and enable a valuable comparison between the two versions, a three-factor structure was selected for this analysis.

Figure 1

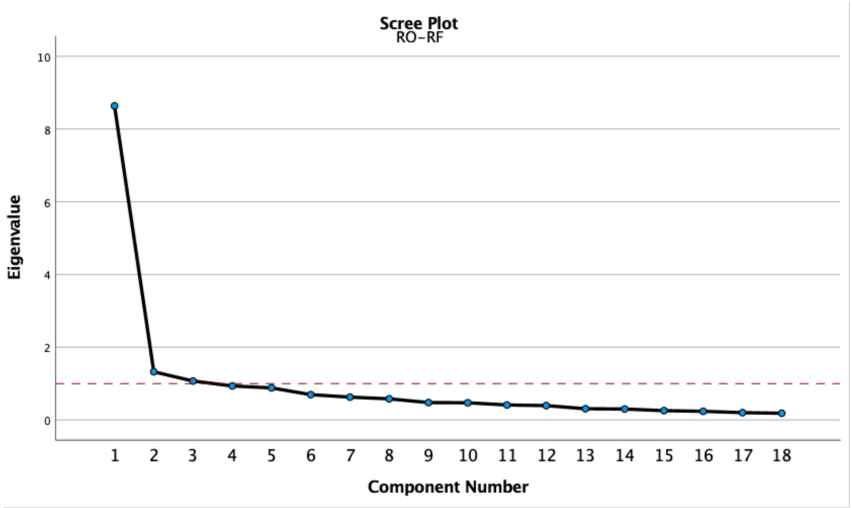
Scree-Plot RQ-OF



Note. The dotted red line indicates an Eigenvalue of one.

Figure 2

Scree-Plot RQ-RF



Note. The dotted red line indicates an Eigenvalue of one.

Table 2 provides an overview of the factor loadings of RQ-OF and RQ-RF after rotation. Based on the rotated factors of the RQ-OF group, it is observable that all EWB items of the RQ-OF group loaded well on the first component as well as four (from the original questionnaire) out of nine (44.44%) SWB items on the second component. It is noteworthy that three of the four added SWB items loaded unexpectedly on the third component. Of the PWB items, only two out of six (33.33%) items loaded on the expected third component.

For the RQ-RF group, all EWB items loaded on one component. Of the nine SWB items, only four items (44.44%) loaded on one component. Of the six PWB items, only three items (50%) loaded on the third component.

Comparing the factor structure of both versions, found in Table 2, the results suggested that the RQ-OF and RQ-RF versions have similar fits on the EWB and SWB scale. However, neither scale demonstrated a good fit to the expected factor structure, with most items of the scales loading on random components. While the RQ-RF scale showed a slightly better fit compared to the RQ-OF scale, both still exhibited poor fit to the expected factor structure. This means that the items in the respective subscales did not align with the expected components but loaded on multiple factors instead. As a result, it can be concluded that the subscales are not well represented in the data of both RQ-OF and RQ-RF.

Table 2

Factor Loadings RQ-OF and RQ-RF After Oblimin Rotation

Item	RQ-OF			RQ-RF		
	Component 1	Component 2	Component 3	Component 1	Component 2	Component 3
1 (EWB)	**.820**	-.038	-.021	.050	**-.874**	-.091
2 (EWB)	**.779**	.026	-.043	.085	**-.763**	.000
3 (EWB)	**.933**	-.079	.009	.078	**-.907**	-.100
4 (SWB)	.124	**.517**	-.084	.012	-.183	**.634**
5 (SWB)	.105	**.827**	.235	-.044	-.108	**.588**
6 (SWB)	.282	.369	-.077	.251	-.127	.202
7 (SWB)	-.141	**.600**	-.263	**.680**	.051	.058
8 (SWB)	.066	**.715**	.038	-.064	-.070	**.676**
9 (PWB)	**.698**	.016	-.091	-.061	**-.729**	.161
10 (PWB)	**.738**	.104	-.009	-.020	**-.715**	.187
11 (PWB)	.113	-.001	**-.692**	**.708**	-.047	.067
12 (PWB)	-.095	**.536**	-.309	.106	.165	**.781**
13 (PWB)	-.241	.157	**-.440**	.193	-.055	**.574**
14 (PWB)	**.615**	-175	-.172	.153	**-.545**	.274
15 (SWB)	.231	.316	-.385	.337	-.039	**.531**
16 (SWB)	.165	-.010	**-.739**	**.763**	-.164	-.073
17 (SWB)	.014	.011	**-.856**	**.872**	.023	.061
18 (SWB)	.030	-.020	**-.859**	**.913**	-.037	-.066

Note. Extraction Method: Principal Component Analysis with oblimin rotation. Factor loadings above .40 are in bold.

RQ2: Internal Consistency

To determine the internal consistency of the RQ-OF and RQ-RF, the reliability on an item and scale level was analysed, using Cronbach's alpha and item-total correlation.

Cronbach's Alpha

At first, the reliability of the RQ-OF and RQ-RF on a total and subscale level were determined using Cronbach's Alpha. On a total scale level, results revealed excellent and highly comparable reliability for both the RQ-OF (α = .93; 95% CI = .916, .933) and the RQ-RF (α = .93; 95% CI = .926, .941). The RQ-OF subscales EWB (α = .87; 95% CI = .850, .884), SWB (α = .85; 95% CI = .834, .867) and PWB (α = .84; 95% CI = .817, .855) also showed high reliability. Similarly, internal consistency was high for the RQ-RF subscales EWB (α =.88; 95% CI = .868, .899), SWB (α = .87; 95% CI = .852, .882) and PWB (α = .85; 95% CI = .828, .864) and highly comparable to the RQ-RF. As the alphas of both versions are similar and the confidence intervals overlap, a test of the difference is not needed.

In conclusion, both RQ-OF and RQ-RF are equal in terms of internal consistency on both a total and a subscale level.

Item-Total Correlations

To assess the factors that contribute the most to the scale's internal consistency, the item-total correlations were computed at both the item and subscale level of the different versions.

First, the item-total correlations for each item with the total scale were analysed, which are displayed in Table 3. The item-total correlations of both versions exceeded .4, indicating good discrimination. Item 5 of the SWB scale indicated the lowest item-total correlation in both versions (RQ-OF r = .471; RQ-RF r = .489) while item 14 of the PWB scale showed the highest score in both versions (RQ-OF r = .761; RQ-RF r = .782).

Next, the item-total correlations were analysed on a subscale level. The results were consistent with those obtained at the item-level, showing high discrimination for all items

within each subscale. In particular, the EWB scale displayed consistently high correlations, with item 3 achieving the highest scores (RQ-OF r = .823; RQ-RF r = .821) and item 2 scoring the lowest (RQ-OF r = .699; RQ-RF r = .719). While the correlations for the SWB scale were lower compared to the EWB items, they were still high, with item 5 scoring the lowest (RQ-OF r = .481; RQ-RF r = .493) and item 17 achieving the best scores (RQ-OF r = .672; RQ-RF r = .776). Finally, within the PWB scale, item 12 achieved the lowest scores (RQ-OF r = .492; RQ-RF r = .492) while item 14 scoring the highest scores (RQ-OF r = .688; RQ-RF r = .740).

Overall, there are no observable differences between the item-total correlations of the RQ-OF and RQ-RF group on total and subscale levels. It can be observed that all items and subscales have demonstrated excellent discrimination, with item-total correlations above .4. Therefore, it can be concluded that both versions possess good discriminative power.

Table 3

Item-Total Correlations on an Item and Subscale level

	RQ-OF				RQ-RF			
	Total Scale	EWB Subscale	SWB Subscale	PWB Subscale	Total Scale	EWB Subscale	SWB Subscale	PWB Subscale
1 (EWB)	.644	.725			.667	.789		
2 (EWB)	.674	.699			.680	.719		
3 (EWB)	.685	.823			.717	.821		
4 (SWB)	.513		.498		.638		.594	
5 (SWB)	.471		.481		.489		.493	
6 (SWB)	.534		.500		.444		.423	
7 (SWB)	.500		.526		.547		.577	
8 (SWB)	.517		.521		.508		.474	
9 (PWB)	.636			.629	.650			.646
10 (PWB)	.671			.660	.700			.688
11 (PWB)	.623			.588	.662			.582
12 (PWB)	.532			.492	.537			.492
13 (PWB)	.642			.644	.633			.649
14 (PWB)	.761			.688	.782			.740
15 (SWB)	.715		.688		.715		.680	
16 (SWB)	.700		.661		.696		.673	
17 (SWB)	.686		.672		.749		.776	
18 (SWB)	.677		.652		.728		.727	

Table 4

Floor and Ceiling Frequencies, Kurtosis (K), Skewness (S), and Chi-Square Test of Differences in Ceiling Effects in RQ-OF and RQ-RF

Item	RQ-OF				RQ-RF				$X^2(df)$	*p*
	0 (n(%))	5 (n(%))	K	*S*	0(n(%))	5(n(%))	K	*S*		
1 (EWB)	7(1)	97(14.3)	0.56	-0.89	7(1.1)	93(14)	-0.04	-0.55		
2 (EWB)	10(1.5)	219(32.2)[a]	1.24[b]	-1.21[b]	5(1.8)	158(23.9)[a]	0.66	-0.76	11.55(1)	<.001
3 (EWB)	11(1.6)	143(21.0)[a]	0.76	-1.06[b]	4(0.6)	136(20.5)[a]	0.02	-0.59	0.05(1)	.827
4 (SWB)	78(11.5)	61(9.0)	-0.99	-0.26	27(4.1)	52(7.9)	-0.55	-0.01		
5 (SWB)	84(12.4)	35(5.1)	-0.92	-0.12	14(2.1)	21(3.2)	-0.08	-0.02		
6 (SWB)	9(1.3)	196(28.8)[a]	0.55	-1.02[b]	2(0.3)	151(22.8)[a]	0.40	-0.57	6.33(1)	.012
7 (SWB)	94(13.8)	106(15.6)[a]	-1.13[b]	-0.30	40(6)	69(10.4)	-0.67	-0.31	7.89(1)	.005
8 (SWB)	37(5.4)	112(16.5)[a]	-0.46	-0.63	6(0.9)	54(8.2)	-0.11	-0.37	21.39(1)	<.001
15 (SWB)	16(2.4)	144(21.2)[a]	-0.01	-0.74	7(1.1)	91(13.7)	-0.24	-0.38	0.12(1)	.733
16 (SWB)	18(2.6)	155(22.8)[a]	-0.01	-0.81	7(1.1)	124(18.7)[a]	-0.04	-0.58	3.36(1)	.067
17 (SWB)	18(2.6)	151(22.2)[a]	0.30	-0.87	9(1.4)	241(15.1)[a]	-0.05	-0.55	11.12(1)	<.001
18 (SWB)	11(1.6)	193(28.4)[a]	0.04	-0.85	10.(1.5)	147(22.2)[a]	-0.02	-0.63	0.56(1)	.453
9 (PWB)	15(2.2)	229(33.7)[a]	0.85	-1.17[b]	7(1.1)	155(23.4)[a]	0.35	-0.78	17.31(1)	<.001
10 (PWB)	22(3.2)	168(24.7)[a]	0.52	-1.04[b]	9(1.4)	114(17.2)[a]	0.34	-0.69	11.32(1)	<.001
11 (PWB)	18(2.6)	183(26.9)[a]	-0.19	-0.78	7(1.1)	138(20.8)[a]	-0.19	-0.61	6.78(1)	.009
12 (PWB)	86(12.6)	74(10.9)	-1.10[b]	-0.16	30(4.5)	40(6)	-0.59	-0.03		
13 (PWB)	18(2.6)	146(21.5)[a]	-0.05	-0.75	9(1.4)	97(14.7)	0.08	-0.48	10.52(1)	.001
14 (PWB)	21(3.1)	174(25.6)[a]	-0.10	-0.84	7(1.1)	116(17.5)[a]	0.07	-0.61	12.88(1)	<.001

Note. Response options for the RQ-OF included (0) never, (1) once or twice a month, (2) about once a week, (3) 2 or 3 times a week, (4) almost every day, (5) every day. The RQ-RF option included (0) never, (1) rarely, (2) sometimes, (3) regularly, (4) often, (5) (almost) always. Chi-square values were calculated if a ceiling or floor effect was indicated.

[a] Ceiling effects

[b] Skewness or Kurtosis

RQ3: Distribution of Items and Scales

Item Distribution

To gain insight into the distribution of the items the skewness, kurtosis, and the ceiling and floor frequencies of the two versions RQ-OF and RQ-RF were examined. Table 4 provides an overview of the skewness, kurtosis, frequency distribution of the floor and ceiling scores and chi-square test of the differences in ceiling scores sorted by scale.

RQ-OF. Results from the skewness and kurtosis measure revealed that items 2, 3, 6, 7, 9, 10, and 12 were only skewed, distorted or both, as the statistics, *S* and κ, were not between the range of +1 and -1. The item frequencies indicated ceiling effects for items 2 and 3 of the EWB scale, items 6, 7, 8, 15, 16, 17, and 18 of the SWB scale, and items 9, 10, 11, 13, and 14 of the PWB scale as 15% or more of the participants answered with the highest answer option "every day" on these items. No floor effects could be detected, as the response distribution for the lowest item "never" was always below 15%.

RQ-RF. The skewness and kurtosis measure indicated that all items had acceptable values between the range of +1 and -1 indicating a normal distribution. Compared to the RQ-OF, the distribution of the RQ-RF items was closer to the normal distribution. Based on the items' frequencies, ceiling effects were detected for the items 2 and 3 of the EWB scale, items 6, 16, 17, and 18 of the SWB scale and items 9, 10, and 11 of the PWB scale. That is, 15% or more of the participants answered with the highest answer option "(almost) always". Additionally, no floor effects were visible as participants answered the lowest answer option "never" in all cases with less than 15%.

Chi-Square. The RQ-RF demonstrated fewer ceiling effects when compared to the RQ-OF, indicating a more balanced distribution of scores and potentially superior performance in capturing a wider range of responses. This is further supported by the chi-square tests as they indicate a significant decrease in the number of highest response options for several items (2, 8, 17, 9, 10, and 14) in the RQ-RF. Thus, by changing the response

format from original to revised, the RQ-RF seems to have a more desirable distribution of response options compared to the RQ-OF.

Scale Distribution

To determine the distribution of the scale scores of the two versions, the measures of skewness and kurtosis were calculated. The results suggest that the majority of scales had values indicating normality as the statistics, *S* and κ, were between +1 and -1. However, there was an exception for the EWB scale of the RQ-OF, where the scores showed a skewed distribution to the right (*S*= -1.03).

The distribution of mean scores on a scale level was analysed to assess the presence of ceiling and floor effects at the scale level. Table 5 displays the frequencies of the mean scale scores for both versions. Histograms of the frequencies can be found in Appendix B.

Two conditions were considered: strict, where scores of exactly 0 and 5 were defined as floor and ceiling, and rounded, where scores ≥ 4.5 were considered as ceiling and scores < 0.5 as floor. Applying the strict cut-off, neither ceiling nor floor effects were detected for both the RQ-OF and RQ-RF versions, suggesting comparable distributions between the two versions.

When considering scores ≥ 4.5 as ceiling and ≤ 0.5 as floor, ceiling effects were detected for the EWB scale in both the RQ-OF and RQ-RF versions, while no floor effects were observed. Table 6 illustrates the frequencies of the mean scale scores, indicating that both versions were similarly distributed in this aspect.

Table 5

Floor and Ceiling Frequencies of Subscale Scores of RQ-OF and RQ-RF (Strict)

Scale	RQ-OF		RQ-RF	
	0 (n(%))	5 (n(%))	0(n(%))	5(n(%))
EWB	3(.4)	66(9.7)	2(.3)	71(10.7)
SWB	1(.1)	10(1.5)	1(.2)	4(.6)
PWB	2(.3)	28(4.1)	3(.5)	19(2.9)

Note. Response options for the RQ-OF included (0) never, (1) once or twice a month, (2) about once a week, (3) 2 or 3 times a week, (4) almost every day, (5) every day. The RQ-RF option included (0) never, (1) rarely, (2) sometimes, (3) regularly, (4) often, (5) (almost) always. S*trict:* values of exactly 0 and 5 are seen as floor and ceiling effects.

[a] Ceiling effects.

Table 6

Floor and Ceiling Frequencies of Subscale Scores of RQ-OF and RQ-RF (Rounded)

Scale	RQ-OF		RQ-RF	
	0 (n(%))	5 (n(%))	0(n(%))	5(n(%))
EWB	3(.8)	118(17.3) [a]	4(.6)	106(16) [a]
SWB	4(.5)	41(6)	3(.5)	32(4.9)
PWB	4(.5)	95(13.9)	4(.7)	65(9.8)

Note. Response options for the RQ-OF included (0) never, (1) once or twice a month, (2) about once a week, (3) 2 or 3 times a week, (4) almost every day, (5) every day. The RQ-RF option included (0) never, (1) rarely, (2) sometimes, (3) regularly, (4) often, (5) (almost) always. *Rounded*: scores ≥ 4.5 are seen as scores at the ceiling and score ≤ 0.5 as floor.

[a] Ceiling effects.

To detect differences between the mean scores, independent sample t-tests were conducted. An overview of the means with standard deviation, t-tests, and effect size of the

two different versions can be found in Table 7. The results revealed no significant differences in mean scores between the EWB, SWB and PWB scales of the different versions RQ-OF and RQ-RF.

Table 7

Mean, Standard Deviation of EWB, SWB, PWB of RQ-OF and RQ-RF and T-Test of the Differences with Effect Sizes of the Differences

Scale	*RQ-OF*		*RQ-RF*		*t(df)*	*p*	*d*
	M	*SD*	*M*	*SD*			
EWB	3.67	1.01	3.57	0.96	-1.99(1340)	.107	-.11
SWB	3.20	0.92	3.21	0.79	0.23(1318)	.821	.12
PWB	3.41	1.00	3.36	0.86	-0.99(1318)	.322	-.05

Note. The Mean Scores were calculated on 6-point scales ranging from 0 to 5.

Discussion

In the past, the Mental Health Continuum-Short Form (MHC-SF) was often used as an instrument to assess mental well-being and its three dimensions according to the tripartite model, namely emotional well-being (EWB), social well-being (SWB) and psychological well-being (PWB, Lamers et al., 2011). Due to several practical issues with the MHC-SF, two revised versions were developed, namely the revised questionnaire with the original response format (RQ-OF) and the revised questionnaire with a revised response format (RQ-RF). This study aimed at comparing the psychometric properties of these two revised versions, RQ-OF and RQ-RF, in terms of their factor structure, distribution of item and scale scores and internal consistency.

The first research question inferred the factor structure of both versions. Results suggested that both the RQ-OF and the RQ-RF version had a similar but poor fit on the

expected three-factor structure. The second research question compared the internal consistency of both versions. The alphas of both versions were similarly high and suggested high reliability. The third research question inferred about the distribution of the versions on both an item and scale level. On an item level, no floor effects were observed in either version. However, in terms of ceiling effects, the RQ-RF demonstrated much fewer ceiling effects and was more normally distributed than the RQ-OF. On a scale level, the RQ-OF and RQ-RF both showed no ceiling effects, nor floor effects and were hence both similarly distributed.

Factor Structure

In the present study, the factor structure of the RQ-OF and RQ-RF was examined and compared in response to the first research question. A three-factor structure was chosen for analysis as suggested from Eigenvalues and as it is consistent with current research (Iasiello et al., 2022). However, although the RQ-RF scale showed slightly better fit compared to the RQ-OF scale, both versions still exhibited poor fit to the expected factor structure. These findings suggest that both versions of the scale were incongruent with the expected three-component model. That is, most items of the scales loaded on unexpected components, indicating poor fit to the expected three-factor structure. Regarding the research question investigating the factor structure of the RQ-OF and RQ-RF, it seems that a three-factor structure does not provide an adequate fit to the data.

Since the RQ-OF and RQ-RF are modified versions of the MHC-SF that have not been assessed before, the results of this study can only be compared to previous studies that examined the factor structure of the MHC-SF. Previous research of the MHC-SF indicated that the instrument showed a good fit for the three-factor structure. For example, a confirmatory factor analysis was performed by Luijten et al. (2019) on the MHC-SF and found a satisfactory to good fit for the tripartite model. This finding has been replicated further by other studies (Joshanloo, 2017; van Zyl & Olckers, 2019). Similarly, Iasiello et al.

(2022) confirmed this result in their systematic review and meta-analysis on the factor structure of the MHC-SF across 26 studies that indicated a good fit for the tripartite model. Thus, the literature on the factor structure of the MHC-SF previously indicated a good fit with the three-factor model. However, following the changes, both revised versions no longer demonstrate a fit with the three-factor model. To gain a deeper understanding of this discrepancy, possible explanations will be explored in the following sections to investigate whether changes in the revised versions led to the inconclusive results.

One explanation for the poor fit of both the RQ-OF and RQ-RF versions compared to the MHC-SF may be that the two versions used in the present study were revised versions from the MHC-SF incorporating additional SWB items and reformulated items. These alterations may have disrupted the anticipated three-factor structure that was originally established as valid in the MHC-SF. In particular, the MHC-SF items were formulated in questions whereas the revised versions were written in statements. Changing the items from questions to statements could have been an issue as it may alter the way participants interpret and respond to the content. Questions typically encourage participants to provide responses based on their personal experiences or beliefs, whereas statements are more authoritative and may elicit a different response pattern. As an example, an individual who previously responded positively to the question "How often did you feel happy in the past month?" as presented in the MHC-SF, may be less inclined to agree with the statement "I am happy," as presented in the revised versions. The reason for this discrepancy could be that the statement "I am happy" might not directly resonate with participants as the previous question did. Instead, they may find a statement like "I felt happy," which describes a momentary emotion, more appropriate, as it conveys a transient feeling rather than a more enduring sense of happiness. Therefore, it seems that the modifications made to the items of the MHC-SF in the revised versions may explain why the three-factor structure no longer exhibits a similar level of fit as observed in the original MHC-SF.

Another possible explanation for the disparity in the fit of the three-factor structure could be the difference in time frames used for response. In the MHC-SF, participants were asked to report their experiences over a one-week period, while in the revised versions, they were asked to reflect on the past month. This discrepancy in time frame may have influenced participants' responses, as individuals may have varying recollections and interpretations of their well-being over longer versus shorter time spans. Hence, the change in time frame could have yielded different outcomes in the revised versions and hence affected the fit of the three-factor structure. Hence, assuming that the MHC-SF initially had a three-factor structure and it changed in the revised versions, the discrepancy in fit could be explained by the mentioned addition of SWB items, reformulation of items, and change in time frame. However, it is important to note that the factor structure of the MHC-SF has been a subject of ongoing debate, suggesting that the original three-factor model may not accurately represent mental well-being.

Indeed, hints of this are provided in the present study as the scree-plot analysis suggested that alternative factor structures, such as a one-factor or two-factor model, could potentially provide a better fit to the data. This is also further supported by the results of the second research question that inferred about the reliability. That is, the results revealed excellent and highly comparable reliability for the versions. As no comparison studies for the revised versions are available, it is necessary to review the literature on the MHC-SF.

Other literature has proposed a range of other factor solutions for the MHC-SF, including a single, two-factor, bi-factor, or hierarchical model (Iasiello et al., 2022). In a study conducted by de Bruin and du Plessis (2015), the three-factor model was compared to a one-factor model and a bi-factor model. Their findings indicated that the traditional three-factor model may not fully capture the components of well-being. Instead, the bifactor model, which includes a general well-being factor and three residualized group factors, demonstrated the best fit for the MHC-SF scale. This suggests the presence of a robust overarching factor of

general well-being, while the specific group factors represent weaker and more specific aspects of well-being.

These findings align with similar results reported in other studies, indicating that the subscales of well-being contribute minimal variance beyond what is explained by the general factor (Santini et al., 2020; Żemojtel-Piotrowska et al., 2018). In light of the research question, these findings suggest that alternative factor structures, specifically the bifactor model, should be considered for RQ-OF and RQ-RF in future research.

To summarize, both the RQ-OF and RQ-RF versions exhibited poor fits for the three-factor structure, leading to inconclusive results regarding their factor structure and hence the first research question. While the literature on the MHC-SF demonstrated a good fit for the three-factor structure, the changes made to the MHC-SF may have affected the factor structure in the revised versions. Additionally, it is possible that a three-factor model might not be sufficient to explain well-being and it is recommended that future research consider alternative factor structures such as one-factor, two-factor, or bi-factor models for the revised versions.

Internal Consistency

For the second research question, the internal consistency was explored by determining the reliability of the two versions on both an item and scale level. Both the RQ-OF and the RQ-RF demonstrated similarly high reliability and internal consistency not only for the whole questionnaires but also for all subscales. The internal consistency of both versions was also very high on the total scale level. This suggests that both versions are reliable instruments that consistently produce similar results for participants across samples and occasions. This indicates that the instruments seem to be good measurement tools as the good reliability increases the confidence in the accuracy and stability of the obtained scores. This finding is comparable to the internal consistency levels ranging from .85 to .88 in previous studies (Keyes & Annas, 2009; Lamers et al., 2011; Luijten et al., 2019).

Interestingly, the alpha value of SWB in those studies has shown to be lower. However, this has not been the case in this study, meaning that the reformulation of items and addition of SWB items in the revised versions used in the present study may have increased the alpha value. This result could be explained by the reformulation or addition of the SWB items. Further research should investigate this phenomenon further.

The RQ-OF and the RQ-RF both exhibit high reliability and internal consistency indicating the versions are precise instruments, even though the underlying factor structure does not seem to be clear. The scree-plots indicate that a one- or two-factor structure may be most suitable for the versions. Furthermore, this result indicates that changing the response format from questions to statements did not affect the overall high internal consistency of both the scale and total level of the questionnaires. This could be due to the fact that the internal consistency of the RQ-OF was already high, as well as saturated, and that a change in response pattern could not result in any visible improvements.

Regarding the second research question, the RQ-OF and RQ-RF versions demonstrate high reliability and internal consistency, suggesting that they are precise measurement instruments. However, more research is needed to examine the underlying factor structure of these versions.

Distribution

The third research question aimed to explore the distribution of scores at the item-level and scale-level, as well as any differences between the RQ-OF and RQ-RF versions. The absence of floor effects was observed for both versions after analyzing the item distributions. However, the RQ-RF showed fewer ceiling effects, indicating that it may be a better measurement tool than the RQ-OF. This suggests that changing the response format led to fewer participants scoring at the highest end of the scale, resulting in a significant improvement in the item distributions. This is a positive outcome since the presence of floor and ceiling effects can hinder the researcher's ability to differentiate between participants who

score at the extreme end of the scale (Terwee et al., 2007). The presence of such effects in the RQ-OF indicates that participants completing that version may have preferred to choose a higher or lower score than what was allowed by the provided answer scale. That is, the highest response option of the RQ-OF, "every day," may not have captured the full range of participants' experiences, leading to a clustering of responses at the upper end.

In contrast, the absence of ceiling effects in the RQ-RF indicates that this version allowed for a greater range of response possibilities. The highest response option, "almost always," may have implicitly encompassed the idea of experiencing something throughout the entire day, while the lower response options allow for distinctions within a day, such as "sometimes during the day" or "rarely." This increased range of response options may have prevented participants from feeling constrained and resulted in a more balanced distribution of responses.

Therefore, it can be concluded that the absence of floor and ceiling effects in the RQ-RF version suggests that this version is able to capture the wide range of responses in a more accurate and reliable way.

The reduction in ceiling effects may be attributed to the participants' improved ability to retrospectively assess the frequency of their feelings within a shorter timeframe of one week, as in the RQ-RF version, compared to the longer timeframe of one month in the RQ-OF version. That is as participants may have difficulty accurately recalling specific instances or frequencies of behaviors or emotions that occurred throughout the entire month. This could lead to a tendency to provide more extreme responses or cluster responses towards the upper end of the scale, resulting in ceiling effects. On the other hand, when participants are asked to reflect on a shorter time frame, such as the last week, it may reduce the demands on memory and improves the accuracy of recall. This is known as the “recency effect” and suggests that participants are more likely to recall events that occurred recently as compared to events that

occurred a while ago (Goldstein, 2011). This may result in a more balanced distribution of responses across the scale and decrease the likelihood of ceiling effects.

On a scale level, the results suggested no difference in distribution, as no ceiling, floor effects and no significant differences in the scale scores were detectable. This means that a change in the response format caused no change in terms of the scale's distribution. This could be explained by the fact that the scales were fairly normally distributed before the change of the response pattern as visible in the histograms of the RQ-OF displayed in the Appendix A. Hence, the change in response pattern could not improve the already acceptable starting situation of the distributions.

Overall, the analysis of the item-level and scale-level distributions of the RQ-OF and RQ-RF versions revealed that the RQ-RF is a more reliable measurement tool than the RQ-OF due to fewer ceiling effects.

Strengths and Weaknesses

One of the key strengths of this study is its large and representative sample size, as the participants were randomly selected from a nationally representative cohort. This randomization ensures that the results are comparable and applicable to the wider Dutch population. Additionally, the demographic characteristics of the two samples, such as gender, marital status, education, and work status, were found to be similar, further increasing the generalizability of the findings.

However, one limitation of the study is that only a limited number of psychometric properties were investigated. For instance, the external validity of the versions has not been assessed. This is an important limitation as it would provide insights into whether the versions accurately measure well-being. Another psychometric property that could have been explored are other types of reliability, such as test-retest reliability, which could provide insights into the stability of the data over time.

Moreover, the factor structure was only examined through exploratory analysis, and future research could use confirmatory factor analysis to gain more insights into the structure of the two versions. Finally, only one subset of the dataset from the different revisions was included in the study, and integrating all analyses from all four revision conditions in future research could provide a more comprehensive understanding of the topic.

Conclusion

The research questions aimed to examine how changing the response format of two revised versions of the MHC-SF affected their psychometric properties. The findings of the present study indicated a reduction in ceiling effects, but no improvement was observed in terms of factor structure, internal consistency, or scale distribution. The poor fit of the three-factor structure of the revised versions raised doubts about its appropriateness. This highlights the importance of further research to explore alternative models, such as a one-factor or bifactor model, to better understand the factor structure of the revised versions. Considering the lack of significant improvements from changing the response format, it is not possible to determine which version of the RQ-OF and RQ-RF is superior in terms of psychometric properties. However, it can be concluded that the changes made to the response format did not negatively affect the psychometric properties of the questionnaire. Consequently, if participants perceive the revised questionnaire or the revised response format as more user-friendly compared to the MHC-SF, they could be implemented in practice. However, the results need to be replicated first to confirm that the revised questionnaire is safe to use.

References

Cattell, R. B. (1966). The Scree Test For The Number Of Factors. *Multivariate Behavioral Research*, *1*(2), 245-276. https://doi.org/10.1207/s15327906mbr0102_10

Chakhssi, F., Kraiss, J. T., Sommers-Spijkerman, M., & Bohlmeijer, E. T. (2018). The effect of positive psychology interventions on well-being and distress in clinical samples with psychiatric or somatic disorders: a systematic review and meta-analysis. *BMC Psychiatry*, *18*(1), 211. https://doi.org/10.1186/s12888-018-1739-2

de Bruin, G. P., & du Plessis, G. A. (2015). Bifactor Analysis of the Mental Health Continuum—Short Form (MHC—SF). *Psychological Reports*, *116*(2), 438-446. https://doi.org/10.2466/03.02.PR0.116k20w6

Fava, G. A., Cosci, F., Guidi, J., & Tomba, E. (2017). Well-being therapy in depression: New insights into the role of psychological well-being in the clinical process. *Depress Anxiety*, *34*(9), 801-808. https://doi.org/10.1002/da.22629

Geisen, E. (2020). *Tips for measuring behavioral frequency.* Retrieved 6. May from https://www.qualtrics.com/blog/measuring-behavioral-frequency/

Goldstein, E. B. (2011). *Cognitive psychology : connecting mind, research, and everyday experience* (3rd ed.). Wadsworth Cengage Learning.

Gotlin, M. J., Kingery, M. T., Baron, S. L., McCafferty, J., Jazrawi, L. M., & Meislin, R. J. (2020). Recall Bias in Retrospective Assessment of Preoperative Patient-Reported American Shoulder and Elbow Surgeons Scores in Arthroscopic Rotator Cuff Repair Surgery. *Am J Sports Med*, *48*(6), 1471-1475. https://doi.org/10.1177/0363546520913491

Huda, A. S. (2021). The medical model and its application in mental health. *International Review of Psychiatry*, *33*(5), 463-470. https://doi.org/10.1080/09540261.2020.1845125

Iasiello, M., Agteren, J., & Muir-Cochrane, E. (2020). Mental Health and/or Mental Illness: A Scoping Review of the Evidence and Implications of the Dual-Continua Model of Mental Health. *Evidence Base*, *2020*. https://doi.org/10.21307/eb-2020-001

Iasiello, M., van Agteren, J., Schotanus-Dijkstra, M., Lo, L., Fassnacht, D. B., & Westerhof, G. J. (2022). Assessing mental wellbeing using the Mental Health Continuum—Short Form: A systematic review and meta-analytic structural equation modelling. *Clinical Psychology: Science and Practice*, *29*, 442-456. https://doi.org/10.1037/cps0000074

IBM Corp. (2021). *IBM SPSS Statistics for Macintosh, Version 28.0.* In IBM Corp.

Jeste, D. V., Palmer, B. W., Rettew, D. C., & Boardman, S. (2015). Positive psychiatry: Its time has come. *The Journal of Clinical Psychiatry*, *76*(6), 675-683. https://doi.org/10.4088/JCP.14nr09599

Joshanloo, M. (2017). Factor Structure and Measurement Invariance of the MHC-SF in the USA. *European Journal of Psychological Assessment*, *35*, 1-5. https://doi.org/10.1027/1015-5759/a000425

Joshanloo, M. (2022). Stability and Change in Subjective, Psychological, and Social Well-Being: A Latent State-Trait Analysis of Mental Health Continuum–Short Form in Korea and the Netherlands. *Journal of Personality Assessment*, 1-9. https://doi.org/10.1080/00223891.2022.2098755

Joshanloo, M., Wissing, M. P., Khumalo, I. P., & Lamers, S. M. A. (2013). Measurement invariance of the Mental Health Continuum-Short Form (MHC-SF) across three cultural groups. *Personality and Individual Differences*, *55*(7), 755-759. https://doi.org/https://doi.org/10.1016/j.paid.2013.06.002

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151. https://doi.org/10.1177/001316446002000116

Keyes, C., & Annas, J. (2009). Feeling good and functioning well: Distinctive concepts in ancient philosophy and contemporary science. *The Journal of Positive Psychology*, *4*, 197-201. https://doi.org/10.1080/17439760902844228

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

Keyes, C. L. M. (2002). The Mental Health Continuum: From Languishing to Flourishing in Life. *Journal of Health and Social Behavior*, *43*(2), 207-222. https://doi.org/10.2307/3090197

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

Keyes, C. L. M. (2006). Subjective well-being in mental health and human development research worldwide: An introduction. *Social Indicators Research*, *77*(1), 1-10. https://doi.org/10.1007/s11205-005-5550-3 (Subjective Well-Being in Mental Health and Human Development Research Worldwide)

Keyes, C. L. M., & Magyar-Moe, J. L. (2003). The measurement and utility of adult subjective well-being. In *Positive psychological assessment: A handbook of models and measures.* (pp. 411-425). American Psychological Association. https://doi.org/10.1037/10612-026

Keyes, C. L. M., & Waterman, M. B. (2003). Dimensions of well-being and mental health in adulthood. In M. H. Bornstein, L. Davidson, C. L. M. Keyes, & K. A. Moore (Eds.), *Well-being: Positive development across the life course.* (pp. 477-497). Lawrence Erlbaum Associates Publishers. http://ezproxy2.utwente.nl/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2003-02621-033&site=ehost-live

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 & Psychotherapy*, *15*(3), 181-192. https://doi.org/10.1002/cpp.572 (Cross-cultural analysis of quality of life and mental health)

Köhle, N. (2010). 'Mag ik vragen, wat ik u moet vragen?': Een onderzoek naar de zwakheden van het Mental Health Continuum Short Form met behulp van het Three-Step Test-Interview. *Bachelor's thesis, University of Twente*.

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. https://doi.org/10.1002/jclp.20741

Leamy, M., Bird, V., Le Boutillier, C., Williams, J., & Slade, M. (2011). Conceptual framework for personal recovery in mental health: systematic review and narrative synthesis. *Br J Psychiatry*, *199*(6), 445-452. https://doi.org/10.1192/bjp.bp.110.083733

Luijten, C. C., Kuppens, S., van de Bongardt, D., & Nieboer, A. P. (2019). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF) in Dutch adolescents. *Health and Quality of Life Outcomes*, *17*(1), 157. https://doi.org/10.1186/s12955-019-1221-y

Martin, E. (2005). Survey Questionnaire Construction. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 723-732). Elsevier. https://doi.org/https://doi.org/10.1016/B0-12-369398-5/00433-3

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*(6), 1069-1081. https://doi.org/10.1037/0022-3514.57.6.1069

Santini, Z. I., Torres-Sahli, M., Hinrichsen, C., Meilstrup, C., Madsen, K. R., Rayce, S. B., Baker, M. M., Ten Have, M., Schotanus-Dijkstra, M., & Koushede, V. (2020). Measuring positive mental health and flourishing in Denmark: validation of the mental health continuum-short form (MHC-SF) and cross-cultural comparison across three countries. *Health and Quality of Life Outcomes*, *18*(1), 297. https://doi.org/10.1186/s12955-020-01546-2

Seligman, M., & Csikszentmihalyi, M. (2000). Positive Psychology: An Introduction. *The American psychologist*, *55*, 5-14. https://doi.org/10.1037/0003-066X.55.1.5

Terwee, C., Bot, S., Boer, M., van der Windt, D., Knol, D., Dekker, J., Bouter, L., & De Vet, H. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of clinical epidemiology*, *60*, 34-42. https://doi.org/10.1016/j.jclinepi.2006.03.012

van Zyl, L. E., & Olckers, C. (2019). The Mental Health Continuum-Short Form in organisational contexts: Factorial validity, invariance, and internal consistency. *European Journal of Mental Health*, *14*, 230-259. https://doi.org/10.5708/EJMH.14.2019.2.2

Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, *64*(4), 678-691. https://doi.org/10.1037/0022-3514.64.4.678

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. https://doi.org/10.1007/s10804-009-9082-y (Positive psychology and adult development)

World Health Organization, P. (2004). *Promoting Mental Health: Concepts, Emerging Evidence, Practice : Summary Report*. World Health Organization. https://books.google.nl/books?id=AxU1RgAACAAJ

Żemojtel-Piotrowska, M., Piotrowski, J. P., Osin, E. N., Cieciuch, J., Adams, B. G., Ardi, R., Bălţătescu, S., Bogomaz, S., Bhomi, A. L., Clinton, A., de Clunie, G. T., Czarna, A. Z., Esteves, C., Gouveia, V., Halik, M. H. J., Hosseini, A., Khachatryan, N., Kamble, S. V., Kawula, A., . . . Maltby, J. (2018). The mental health continuum-short form: The structure and application for cross-cultural studies–A 38 nation study [https://doi.org/10.1002/jclp.22570]. *Journal of Clinical Psychology*, *74*(6), 1034-1052. https://doi.org/https://doi.org/10.1002/jclp.22570

Appendix A

Questionnaire Items

Items in Dutch: Used for Data Collection

De volgende vragen beschrijven gevoelens die mensen kunnen hebben. Lees iedere uitspraak zorgvuldig door en kies het antwoord dat het best weergeeft HOE VAAK u dat gevoel had GEDURENDE DE AFGELOPEN MAAND (RQ-OF)/WEEK (RQ-RF)?

In de afgelopen maand/week, hoe vaak had u de volgende gevoelens?

1. Ik ben gelukkig.

2. Ik ben geïnteresseerd in het leven.

3. Ik ben tevreden met mijn leven.

4. Ik doe iets waardevols voor onze samenleving.

5. Ik denk dat ons land zich goed ontwikkelt.

6. Ik accepteer anderen zoals ze zijn.

7. Ik hoor bij een groep mensen, mijn buurt of stad.

8. Ik begrijp hoe onze samenleving werkt.

9. Ik accepteer mezelf zoals ik ben.

10. Ik heb grip op mijn leven.

11. Ik deel lief en leed met enkele mensen.

12. Ik word uitgedaagd om te groeien.

13. Ik durf mijn ideeën te uiten.

14. Ik heb het gevoel dat mijn leven zin heeft.

15. Ik kan iets betekenen voor anderen.

16. Ik ben tevreden met mijn sociale contacten.

17. Ik voel me verbonden met andere mensen.

18. Ik kan bij andere mensen terecht.

Items in English: Used in this Analysis

The following questions describe feelings people may have. Read each statement carefully and choose the answer that best represents how often you had that feeling DURING THE LAST MONTH (RQ-OF)/WEEK (RQ-RF)

In the past month/week, how often did you have the following feelings?

1. I am happy.

2. I am interested in life.

3. I am satisfied with my life.

4. I am doing something valuable for our society.

5. I think our country is developing well.

6. I accept others as they are.

7. I belong to a group of people, my neighborhood or town.

8. I understand how our society works.

9. I accept myself as I am.

10. I have a grip on my life.

11. I share joys and sorrows with some people.

12. I am challenged to grow.

13. I dare to express my ideas.

14. I feel that my life has meaning.

15. I can mean something to others.

16. I am satisfied with my social contacts.

17. I feel connected to other people.

18. I can reach out to other people.

Response Format RQ-OF

Response Options in Dutch**: **Used for Data Collection

0. Nooit

1. Eén of twee keer per maand
2. Ongeveer 1 keer per week
3. 2 of 3 keer per week
4. Bijna elke dag
5. Elke dag

Response Options in English: Used in this Analysis

0. Not at all
1. Once or twice a month
2. About once a week
3. 2 or 3 times a week
4. Almost every day
5. Everyday

Response Format RQ-RF

Response Options in Dutch: Used for Data Collection

0. Nooit
1. Zelden
2. Soms
3. Regelmatig
4. Vaak
5. (Bijna) altijd

Response Options in English: Used in this Analysis

0. Not at all
1. Rarely
2. Sometimes
3. Regularly
4. Often

5. (Almost) all the time

Appendix B

Distribution of Participant Responses

Table B1

Histogram of Participant Responses per Scale