

The Psychometric Properties of the Forms of Self-Criticising/Attacking and Self-Reassuring Scale (FSCRS) Across Eating Disorders

Kira Lanze

Faculty of Behavioural, Management and Social Sciences
Department of Psychology, Health & Technology
University of Twente

Master Thesis

1st Supervisor: Sander de Vos

2nd Supervisor: Erik Taal

February 2023

Abstract

Background: Several studies demonstrated self-criticism being an independent and robust predictor of eating disorder (ED) pathology. Similarly, self-reassuring strategies buffer against the effects of self-critical attitudes and may increase wellbeing, eventually facilitating treatment success. Hence, management of self-criticism should be included in treatment and monitored. One well-validated tool, frequently used in ED research, is the Forms of Self-Criticising/Attacking and Self-Reassuring Scale (FSCRS). As ED diagnoses (Anorexia Nervosa, Bulimia Nervosa, Binge Eating Disorder, Other Specified Feeding or Eating Disorders) have distinct symptom presentations, it is unclear whether self-criticism is measured the same across these types. Further investigation of the FSCRS in this regard is required to ensure its valid applicability in practice. Thus, the psychometric properties of the FSCRS between the named ED types was examined in depth, by determining factor structure, measurement invariance across the named ED types and convergent validity with mental wellbeing.

Methods: The conducted research was a cross-sectional, quantitative study. Measures were taken from an existing dataset, collected at Human Concern Centre for Eating Disorders in the Netherlands. 981 responses could be analysed. Used materials included the FSCRS, Mental Health-Continuum Short Form, and Eating Disorder Examination Questionnaire.

Results: Confirmatory Factor Analysis (CFA) revealed a three-factor solution being superior to a two-factor solution. Also, multigroup CFA identified measurement invariance on three levels. Not enough Pearson's correlations were strong enough to be considered convergent validity. Nonetheless, F-tests revealed no significant differences in correlations between ED types.

Conclusion: Multigroup CFA showed that different ED patients conceptualised self-criticism the same way, which allows for a generalised use of this scale in ED research and practice. Still, the results just met the established criteria and are the first in this specific context. Convergent validity could not be identified for the whole scale. Yet, self-reassurance showed adequate convergent validity. Most correlations were also adequate to strong, and findings are in line with prior research. Hence, if not such strict criteria were used, convergent validity probably could have been identified. As no differences between ED types were found, the FSCRS is interpreted as valid for the different EDs in this regard.

Keywords: self-criticism, self-reassurance, eating disorders, FSCRS, MHC-SF, cross-sectional

Table of Contents

Introduction	4
Self-Criticism.....	4
Eating Disorders	4
Self-Criticism in Eating Disorders	5
The Forms of Self-Criticising/Attacking and Self-reassuring Scale (FSCRS).....	6
Psychometric Properties of the FSCRS	7
The Present Research.....	8
Methods.....	9
Design and Procedure.....	9
Participants	10
Materials	10
The Eating Disorder Examination Questionnaire (EDE-Q)	10
The Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale (FSCRS).....	11
Mental Health Continuum – Short Form (MHC-FS).....	11
Data Analysis.....	11
Confirmatory Factor Analysis.....	12
Measurement Invariance	13
Convergent Validity.....	14
Results	14
Description of the Sample	14
Hypothesis 1: Confirmatory Factor Analysis	16
Hypothesis 2: Measurement Invariance with Multigroup Confirmatory Factor Analysis ...	17
CFA per ED	18
Configural Model Test.....	18
Metric Model Test.....	18
Scalar Model Test	18
Hypothesis 3: Convergent Validity Total Sample.....	20

Hypothesis 4: Convergent Validity between Groups	22
Discussion	22
Theoretical Reflection	23
Hypothesis 1.....	23
Hypothesis 2.....	23
Hypothesis 3 and 4.....	24
Strengths and Limitations.....	26
Implications for Practice and Future Research.....	27
References	28
Appendix A: The Eating Disorder Examination Questionnaire.....	34
Appendix B: Forms of Self-Criticizing/ - Attacking and Self-reassuring Scale Questions	34
Appendix C: Mental Health Continuum Short-Form Questionnaire	34
Appendix D: Table with Standardised Factor Loadings for the Two-Factor Solution	35
Appendix E: Partial Measurement Invariance	36

This thesis aims at validating a questionnaire on self-criticism and self-reassurance in the context of eating disorders (EDs). Starting point is an introduction of self-criticism and its role in EDs, which is investigated further, by reviewing EDs in more detail. At last, the questionnaire will be introduced and current knowledge on its psychometric properties are displayed. Based on this, it will be examined whether self-criticism is conceptualised the same way among individuals with different ED types and if self-criticism measured by this instrument is also related to mental wellbeing.

Self-Criticism

Although there is no clear definition of self-criticism, this construct is usually conceptualised according to an excessive focus on achievement and overly harsh and critical evaluation of the self as well as performance (e.g., Kelly & Carter, 2013; Löw et al., 2020). Used as a non-functional emotion regulation strategy, physical and/or mental self-attacks or punishments shall correct a certain behaviour, avoid mistakes, or escape negative feelings after perceived failure (Gilbert et al., 2004; Perkins et al., 2020). Specifically, Gilbert and colleagues (2004) further differentiated between two types of self-criticising. One centring around dwelling on mistakes, including a sense of inadequacy, while the second deals with wanting to hurt oneself and feelings of self-hatred and -disgust (Gilbert et al., 2004).

Perceived negative outcomes are usually solely related to one's own deficits (Noordenbos et al., 2014). Vulnerable feelings, such as pain or fear, are likely to be labelled as unacceptable from early age on and are later misinterpreted as shameful or wrong. Transferred to the perception of oneself, a set of self-critical shame-based core beliefs is created, oftentimes central to psychological disorders and especially observable in EDs (Fennig et al., 2008; Munro et al., 2017).

On the other hand, self-reassurance (i.e., being able to sooth and pacify the self when things go wrong; Gilbert et al., 2004) might function as a potential buffer against self-critical attitudes and its effects (Sommers-Spijkerman et al., 2018). In that sense, researchers found self-reassuring strategies to predict a protective and caring relationship with one's own body and decreased severity of eating psychopathology symptoms (Mendes et al., 2019; Duarte, et al., 2017). Moreover, self-reassurance could be associated with higher wellbeing when attempting for weight management which is likely to facilitate treatment success (Duarte et al., 2019). On the contrary, as already implied, self-criticism itself is linked to the onset as well as maintenance of EDs.

Eating Disorders

The fifth version of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V) by the American Psychiatric Association (APA) defines EDs as being characterised “by a

persistent disturbance of eating or eating-related behavior that results in the altered consumption or absorption of food and that significantly impairs physical health or psychosocial functioning” (APA, 2013; p.329). EDs are highly prevalent in individuals worldwide. A literature review from 2019 revealed a lifetime prevalence of 8.4% in women and 2.2% in men, as well as around 2.2% in Europe generally (Galmiche, et al., 2019). Four main ED types can be distinguished.

Anorexia Nervosa (AN) is marked by a refusal to maintain a normal body weight, resulting in significantly low Body-Mass-Index (BMI; APA, 2013). It can be distinguished into two subtypes, namely restricting- and binge-eating/purging type (APA, 2013). The restricting type is characterised by food avoidance, whereas the latter is characterised by periods of binge eating and compensative purging behaviour (APA, 2013; Bulik et al., 2005).

Next, Bulimia Nervosa (BN) is generally characterised by frequent episodes of binge eating, which are then followed by inappropriate compensatory behaviours, in order to avoid gaining weight (APA, 2013; Williams & Levinson, 2022). These binge eating episodes express themselves through discrete periods of consuming large amounts of food while being/feeling out of control (Cooper & Fairburn, 2003; Williams & Levinson, 2022). Followed by feelings of guilt and shame (Serpell et al., 2020), compensatory behaviours include self-induced vomiting, misuse of laxatives or excessive exercise (APA, 2013).

Like in BN, individuals with Binge Eating Disorder (BED) experience recurrent episodes of binge eating but without regular compensatory behaviours (APA, 2013; Cooper & Fairburn, 2003). Although there is a reliable association with obesity/overweight, obesity is not necessarily associated with disordered eating behaviours (APA, 2013).

Finally, individuals are diagnosed with an Other Specified Feeding or Eating Disorder (OSFED) when ED-typical symptoms are causing clinically significant distress and impair daily functioning but do not meet the full criteria for any of the EDs (APA, 2013), for instance atypical AN, purging disorder or night eating syndrome (Riescoa et al., 2018). Although not very well-known by the general public, OSFED patients made up the vast majority of ED cases world-wide (Galmiche et al., 2019).

Self-Criticism in Eating Disorders

A study by Fennig and colleagues (2008) demonstrated the significant relationship between self-criticism and EDs. Also demonstrated by other studies, self-criticism posed an independent and robust predictor of the disorder. Highly self-critical patients showed more severe ED pathology compared to less self-critical individuals (Noordenbos et al., 2014). Thereby self-criticism influences ED core cognitions, such as body dissatisfaction (Perkins et

al., 2020). To illustrate, based on personality traits present in self-critical individuals and ED patients (i.e., perfectionism and resulting obsessionality), Fennig et al. (2008) concluded that self-criticism serves as a specific risk factor for starvation in AN. Moreover, researchers related excessive thoughts on food, urges to eat and purge to BED patients' levels of self-criticism and emotional abuse experiences (Dunkley et al., 2010; Serpell et al., 2020). In terms of self-criticism level differences between ED types, studies found contradicting results. Such as individuals with BN having higher self-criticism levels (Fennig et al., 2008), whereas Duarte and colleagues (2016) found no difference between ED types. Regardless, Williams and Levinson (2022) concluded that it might be possible that the function of self-criticism is similar among the disorder types.

In practice, self-criticism is often used as a form of punishment after breaking made-up food rules (Perkins et al., 2020; Williams & Levinson, 2022). Thereby, restrictive eating and excessive exercise might establish a feeling of pride which shall distract from body dissatisfaction (Kelly & Carter, 2013; Löw et al., 2020). Similarly, bingeing and purging are also likely to be aimed at creating some kind of relief of such unpleasant feelings. Yet, this relief of unpleasantness is only temporary and can even intensify (Kelly & Carter, 2013). In a qualitative study, ED patients described immense struggles to separate from their inner self-critical perspective resulting in more severe ED pathology (Fennig et al., 2008; Brennan et al., 2014). Consequently, in order to successfully treat EDs, the management of self-criticism should be included in treatment and therefore needs to be monitored (Fennig et al., 2008; Baião et al., 2015). Hence, one tool, also frequently used in ED research (e.g., Noordenbos et al., 2014; Duarte et al., 2016), to assess self-criticism is the Forms of Self-Criticising/Attacking and Self-Reassuring Scale, which will be introduced in the following in more detail.

The Forms of Self-Criticising/Attacking and Self-reassuring Scale (FSCRS)

Derived from clinical work with depressed individuals by Gilbert and his colleagues (2004), the *Forms of Self-Criticising/Attacking and Self-Reassuring Scale (FSCRS)* aims at exploring the different ways people treat themselves in the face of setbacks and (perceived) failure, or in particular self-critical tendencies (Baião et al., 2015). The scale assesses the two components of self-criticism (i.e., Inadequate Self (IS) and Hated Self (HS)) as well as self-reassurance (Reassured-Self, RS; Gilbert et al., 2004). These subscales are seen as, although correlated, distinct from each other (Baião et al., 2015; Bellur, et al., 2021; Halamová et al., 2018).

Psychometric Properties of the FSCRS

In the development study, internal consistency, evaluated with Cronbach's alpha, showed to be excellent for all three subscales (Gilbert et al., 2004), which was also replicated by later studies. Assessed populations included mixed clinical as well as non-clinical samples (e.g., Baião et al., 2015; Sommers-Spijkerman et al., 2018). Good test-retest reliability was demonstrated in non-clinical samples and patients diagnosed with, anxiety- and personality disorders (e.g., Castilho, Pinto-Gouveia & Duarte, 2015; Sommers-Spijkerman et al., 2018).

Principal-component analysis specified a three-factor solution in the development study of the scale (Gilbert et al., 2004). Due to the strongly correlating components, some studies successfully combined the two self-criticism subscales into one factor, aiming for a two-factor solution among non-clinical samples (e.g., Halamová et al., 2018). Nevertheless, the initial results of a three-factor solution by Gilbert et al. (2004) could be replicated by a number of other studies using confirmatory factor analyses (Baião et al., 2015; Navarrete et al., 2021; Bellur et al., 2021). Also, Becker (2018) confirmed that the originally proposed factor structure showed to have the best fit in a Dutch ED sample.

Furthermore, measurement invariance, meaning examining whether an instrument possesses the same psychometric properties across heterogeneous groups in order to ensure the same construct has been measured (Chen, 2007), was assessed in mainly clinically depressed and non-clinical samples (e.g., Baião et al., 2015; Bellur et al., 2021). Predominantly done by examining factor structure among groups, the three-factor solution showed good fit for both samples, indicating that the samples conceptualised self-criticism and self-reassurance similarly and the FSCRS was able to detect these constructs in both samples equally well (Bellur et al., 2021). In terms of EDs, there is arguably scarce research on this matter. Especially measurement invariance examinations between the varying ED diagnoses does not exist to this point to the authors knowledge.

Moreover, the FSCRS was analysed in terms of convergent validity. As a subtype of construct validity, convergent validity aims at verifying whether the results of an instrument are coherent in relation to other theoretically related measures (Abma et al., 2016). Thereby, constructs that are related should show strong, significant correlations, which can be interpreted as convergent validity. As a result, Castilho et al. (2015) found all three factors of the FSCRS correlating significantly with psychopathology, such as depression, anxiety, or social dysfunction. The two self-criticism components showed moderate to strong positive correlations, whereas self-reassurance showed a moderate negative correlation. On the other hand, positive constructs, such as self-kindness or common humanity, showed the opposite

results for the three factors, here RS showed a strong correlation. Also, Sommers-Spijkerman and colleagues (2018) demonstrated adequate convergent validity of the FSCRS in a non-clinical sample. Similarly, IS and HS indicated significant, negative correlations with constructs associated with mental wellbeing. Again, similar results were obtained in an ED sample (Becker, 2018).

The Present Research

Concluding, self-criticism plays a significant facilitating role in ED psychopathology and hence poor mental wellbeing (Fennig et al., 2008; Duarte et al., 2017; Mendes et al., 2019). This pathology-facilitating role has been found in all main ED types, where it is assumed that its function, namely alleviating distress stemming from body-dissatisfaction, is likely to be the same (Kelly & Carter, 2013; Duarte et al., 2016; Löw et al., 2020; Williams & Levinson, 2022). Consequently, self-criticism plays an important role in later treatment which poses the need for a reliable and valid measurement device for, for instance, clinicians (Fennig et al., 2008; Growth-Marnat & Wright, 2016). Notably, EDs are different diagnoses with distinct symptom presentations. It is not clear whether the FSCRS measures the same construct across these ED types. Thus, further investigation of the FSCRS in this regard is required to ensure its valid applicability in practice (Growth-Marnat & Wright, 2016). Therefore, the psychometric properties of the FSCRS between the different ED types (i.e., AN, BN, BED, OSFED) should be examined in depth. Hence, this study aims at adding to the existing knowledge about the psychometric properties of the full version of the FSCRS in ED patients, by examining factor structure, measurement invariance and convergent validity. To do so, several research questions (RQ) as well as hypotheses (H) were formulated:

RQ1: Does a three-factor structure show a superior fit over a two-factor structure in the FSCRS in an ED patient sample?

H1: It is hypothesised that a three-factor solution (meaning IS, HS, and RS) will provide the better fit compared to a two-factor solution (self-criticism and RS) in this study as prior research has already demonstrated this in various contexts (e.g., Gilbert et al., 2004; Navarrete et al., 2021).

RQ2: Does the conceptualisation of self-criticism and self-reassurance differ across the different ED types?

H2: It is likely that there will be invariance as self-criticism has shown to inherit a similar function among different ED types (Williams & Levinson, 2022), which will probably result in a similar conceptualisation of such (Bellur et al., 2021).

RQ3: In how far can convergent validity between measures of the FSCRS and the construct of mental wellbeing be identified in people with EDs?

H3: As self-criticism appears to have a diminishing effect on mental wellbeing, a strong negative correlation is expected. On the other hand, as self-reassurance seems to enhance mental wellbeing, a strong positive correlation can be expected (Duarte et al., 2017; Mendes et al., 2019).

RQ4: In case of convergent validity, are there differences in how strong FSCRS measures and mental wellbeing measures correlate between ED types?

H4: It is hypothesised that the three FSCRS components will correlate similarly with mental wellbeing among all four ED types, since patients with these diagnoses show similar core ED cognitions and resulting functions of self-criticism (Perkins et al., 2020; Williams & Levinson, 2022).

Methods

Design and Procedure

This instrument validation study is of quantitative nature and has a cross-sectional design. Measures were taken from an existing dataset, which was collected at Human Concern Centre for Eating Disorders in the Netherlands.

Participants were recruited in a timespan from January 2016 to January 2020. The Human Concern Centre offers specialist treatment to men and women who were diagnosed with an ED. The main criteria for treatment and thus participation in the study were that participants (1) had no other main diagnosis apart from an ED, (2) showed no active suicidal actions, (3) did not suffer from severe somatic parameters, and (4) were at least 17 years of age.

Diagnoses were obtained by a psychiatrist in collaboration with a multi-disciplinary intake team. Via semi-structured interviews, demographic as well as psychopathological information of the participants were gathered. Recorded were the participants' age, gender, primary diagnosis, and body mass index (BMI; kg/m²). Participants were briefed about the research's purpose and were asked to give informed consent before actual participation. This consent form explained that participants could refrain from the study at any given moment if

desired. Participants gave their consent to analyse their anonymised data in this study. Finally, the study was ethically approved by the Behavioural, Management and Social Sciences Ethics Committee of the University of Twente.

Participants

Originally, 1176 participants were recruited for this study. After excluding incomplete cases (195), including those who did not sign the informed consent, 981 participants diagnosed with one of the four described EDs remained. For an overview and a more detailed description of participant characteristics, see Table 1 in the results section.

Materials

The following materials were used: apart from a questionnaire about general demographics collected during intake, three questionnaires were given to the participants. These measured patients' ED psychopathology, self-criticism, and mental wellbeing.

The Eating Disorder Examination Questionnaire (EDE-Q)

Developed by Fairburn and Beglin (1994), the Eating Disorder Examination Questionnaire (EDE-Q) is a self-report measurement which aims at assessing eating behaviours in terms of severity and frequency (APA, 2013; Appendix A). Translated into Dutch by Van Furth (2000), participants indicated their answers on a 7-point Likert scale, ranging from 0 ('not one day') to 7 ('every day') in a timespan of four weeks (Aardoom, et al., 2012). From the 36 items, 22 items assess the core attitudinal features of EDs, which were used in this study (Fairburn & Beglin, 1994; Aardoom, et al., 2012). The global EDE-Q score consists of four subscales of ED pathology. These include restraint (5 items, example item: 'Have you been deliberately trying to limit the amount of food you eat to influence your shape or weight?'), eating concern (5 items, example item: 'Has thinking about food, eating or calories made it very difficult to concentrate on things you are interested in?'), shape concern (8 items, example item: 'Have you felt fat?'), and weight concern (5 items, example item: 'How dissatisfied have you been with your weight?'). To obtain the global score, subscale items are summed and averaged. A higher score implies a more severe ED pathology (Aardoom, et al., 2012). Previous research indicated strong reliability with Cronbach alphas ranging from .81 to .91 for the four subscales and a global score of .95 (Aardoom, et al., 2012). Results of this study were able to identify good reliability, however as this questionnaire was mainly used for descriptive purposes, only the reliability of the global score was assessed ($\alpha_{global} = .85$).

The Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale (FSCRS)

To measure self-criticism, the Dutch version of the *Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale (FSCRS)* was used (Appendix B). The scale was developed by Gilbert and colleagues (2004) and translated by Sommers-Spijkerman et al. (2018). From the original 24 items two were removed due to before assessed poor psychometric properties (Gilbert et al., 2004). Answers can be given on a 5-point Likert scale, ranging from ‘not at all like me’ (0) to ‘extremely like me’ (4) (Kelly & Carter, 2013). The scale can be differentiated into three subscales (Gilbert et al., 2004). Namely, Inadequate Self (IS; 9 items, e.g., ‘I am easily disappointed with myself’), Hated Self (HS; 5 items, e.g., ‘I have a sense of disgust with myself’) and the Reassured Self (RS; 8 items, e.g., ‘I am gentle and supportive with myself’). Scale scores are calculated by mean scores, whereby higher scores imply either higher self-criticism or -reassurance. Previous research in a Dutch sample indicated good internal consistency with a Cronbach’s alpha of at least .80 per subscale ($\alpha_{IS} = .86$, $\alpha_{HS} = .80$ and $\alpha_{RS} = .82$; Sommers-Spijkerman et al., 2018). The present study’s results are in line with previous research ($\alpha_{IS} = .84$, $\alpha_{HS} = .83$ and $\alpha_{RS} = .83$).

Mental Health Continuum – Short Form (MHC-FS)

Secondly, the *Mental Health Continuum – Short Form (MHC-FS)* by Keyes (2002) was filled in by the participants in order to assess their mental wellbeing (Appendix C). The questionnaire was translated by Lamers and colleagues (2011) and consists of 14 items. Answers can be given on a 6-Point Likert Scale, ranging from 0 (‘never’) to 5 (‘every day’); better wellbeing is indicated by higher scores. Items assess how participants experienced varying factors of mental wellbeing during the past month. Hereby, the items can be categorised into three subscales which can all be understood as underlying symptoms of wellbeing domains (Keyes, 2002). These domains include emotional- (3 Items, e.g., ‘...happy?’), social- (5 Items, e.g., ‘...that the way society works makes sense to you?’), and psychological wellbeing (6 Items, e.g., ‘...that you had experiences that challenged you to grow and become a better person?’). Scale scores are calculated by mean scores; whereby higher scores imply higher mental wellbeing. Validity and internal consistency of the Dutch MHC-SF in previous research showed to be good for all three dimensions ($\alpha_{emotional} = .83$, $\alpha_{social} = .74$ and $\alpha_{psychological} = .83$; Lamers et al., 2011). The present research could replicate these findings ($\alpha_{emotional} = .84$, $\alpha_{social} = .73$ and $\alpha_{psychological} = .83$).

Data Analysis

Descriptive statistics were determined via SPSS version 28 (IBM, 2021). Main analyses were conducted using R and SPSS version 28.

First, to get an overview of the sample, descriptive statistics were conducted. Demographic information is displayed in Table 1. Furthermore, Table 1 offers an overview of participants' ED pathology, self-criticism, and mental wellbeing levels.

Following that, the data was checked for the assumption of normality. To do so, the Shapiro-Wilk test was conducted. If the test is significant ($p < .05$), it can be assumed that the data's distribution is non-normal (Ghasemi & Zahediasl, 2012). In this case, all Shapiro-Wilk test results were significant, implying a non-normal data distribution. However, as stated in Ghasemi and Zahediasl (2012), significant results often occur in large sample sizes even at small deviations. Nonetheless, these are likely to have no effect on later parametric test results. Furthermore, QQ-plots were created to draw visually based conclusions on the normality of the data set. The created QQ-plots indicated a widely normal distribution of the sample for all three subscales of the FSCRS and MHC-SF.

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) was conducted in R using the 'lavaan' package (Rosseel, 2012). Compared was the originally established three-factor structure (IS, HS, RS; Gilbert et al., 2004) against the by Halamová et al. (2018) suggested two-factor structure (Self-Criticism, RS), whereby RS inherited the same items in each model. Furthermore, for the two-factor structure, the items of HS and IS were merged. Factor loadings are interpreted as meaningful when exceeding the threshold of .4 (Floyd & Widaman, 1995). For each of the proposed models, goodness-of-fit was evaluated using the Comparative Fit Index (CFI) and the Tucker Lewis Index (TLI). Here, values above .90 indicate good fit (Pirralha, 2020). Next to that, the root-mean-square error of approximation (RMSEA) with a confidence interval of 90 % and standardized root-mean-square residual (SRMR) were evaluated. Values below .08 represent good fit and below .05 excellent fit (Pirralha, 2020). Finally, as it is a popular index whether a model possesses a good fit, Chi-Square tests of model fit were considered as well. Thereby, it is aimed for a non-significant statistic (Chen, 2007). As this inherits a nested structure, a Chi-Squared Difference Test was used to assess which model better fits the observed data. Hereby, significant results imply a significant difference of fit between competing models (Werner & Schermelleh-Engel, 2010). However, when interpreting a Chi-Square test, it needs to be considered that bigger sample sizes automatically inherit bigger test strength than smaller samples. Meaning, even minor deviances can be detected, resulting in a significant Chi-Square test and hence seemingly insufficient model fit indices (Eid, Gollwitzer, & Schmidt, 2017; Chen, 2007). Therefore, as this study inherits a rather big sample size, Chi-Square tests will be interpreted with caution.

Measurement Invariance

Measurement invariance was tested by conducting multigroup confirmatory factor analysis (MGCFA). For MGCFA, the before determined best fitting model was used. Essentially, MGCFA consists of a series of model comparisons defining more and more stringent equality constraints (Hirschfeld & von Brachel, 2014). Hence, in order to conduct MGCFA, the necessary steps described by Pirralha (2020) were followed, including analysis on a configural, metric, and scalar level. Therefore, before the actual analysis, CFAs in each group were run to detect any large deviations. In case of unsatisfactory fit, modification indices were calculated, and error correlations subsequently allowed.

Configural Level. After, the actual MGCFA analysis was conducted by starting with a baseline model. In each group, the same item must be associated with the same factor (Chen, 2007). However, loadings themselves, intercepts and variances may vary (Hirschfeld & von Brachel, 2014). Thereby, this level indicates that latent constructs are similar in each group but not identical (Chen, 2007). Configural invariance exists if this baseline model shows a good fit, meaning it meets the afore determined criteria (Pirralha, 2020).

Metric Level. The second level, a weak-invariance model (metric model) is tested at a factor loading level (Chen, 2007). Factor loadings are fixed to the data and thus constrained to be equally fitted (Pirralha, 2020). Hereby, factor loadings represent the strength of the linear relationship between a factor and its associated items. When the loading of each item on their underlying factor is equal in each ED type, it is likely that the underlying factor has the same unit or interval (Chen, 2007). According to Hirschfeld and von Brachel (2014), weak invariance exists if the fit of the metric invariance is not substantially worse than the fit of the configural model. Following Chen (2007), the cut-off point for substantial decrease in the metric model fit will be $-.01$ for CFI, $.015$ for RMSEA and $.30$ for SRMR. Therefore, if values of the metric model do not exceed these cut-off points, invariance on a metric level can be assumed.

Scalar Level. Finally, a strong invariance-model is fitted to the data. In addition to factor loadings, also item intercepts are constrained to be equal across groups (Pirralha, 2020). Intercepts represent the scale's origin and thus scalar invariance indicates that scores from different groups (ED types) inherit the same unit of measurement, or factor loading, and origin/intercept (Chen, 2007). As before, strong invariance is present if the scalar model fit is not significantly worse than the prior model's fit (Hirschfeld & von Brachel, 2014). For scalar invariance, differences of $> -.01$ for CFI and $>.015$ for RMSEA and $>.01$ for SRMR would indicate non-invariance (Chen, 2007). Differences between models was assessed by using the Chi-Squared Difference Tests.

Convergent Validity

Convergent validity of the FCSRS was determined by calculating bivariate Pearson's correlation coefficients between the subscale scores of the FCSRS and MHC-SF in SPSS version 28 (IBM, 2021). Values ranging from .10 to .30 were considered as weak, between .30 and .50 as moderate, and values larger than .50 as strong correlation (Cohen, 1988). Convergent validity for a correlation can be considered if it is $>.50$. Although exact values can be seen as arbitrary, they provide guidance when to consider convergent validity (Abma, et al., 2016). Added to that, convergent validity is usually identified when 75% or more of correlations meet the cut-off point of .50 (Abma et al., 2016).

By using a general linear model, potential differences in correlations between ED types will be examined. Thereby, a significant F-value ($p < .05$) indicates a significant difference between groups (Sutlive & Ulrich, 1998).

Results

Description of the Sample

The sample had a higher female (97.2%) to male ratio, whereby most participants belonged to the age group of 17 to 30 years. OSFED made up the majority of ED diagnoses. The measured average BMI of all participants was at 22.69 ($SD = 7.63$). Added to that, participants showed on average heightened ED pathology ($M = 3.91$; $SD = 1.16$).

Looking at ED types, especially the distribution of genders across the types appears very equal but Chi-Square Difference test still indicated significant differences in this distribution. In regard to the participants age groups, in relation, AN has the youngest participants, whereas most older participants can be found among those with BED. Generally, younger people made up the majority of patients among the groups. Similarly, patients with AN on average had the lowest BMI and BED patients the highest. BN and OSFED showed nearly identical BMIs. On average, ED pathology intensity appears to be similarly across groups. Also, self-criticism measures seem to be similar across ED types, with the IS subscale being the most prominent. Nonetheless, OSFED patients showed to have lower self-criticism measures than the other three types but also highest RS results. For mental wellbeing there are again only slight differences all centring around lower wellbeing. Still, it needs to be noted that F-Tests still imply significant differences among ED types in terms of ED pathology, self-criticism and mental wellbeing (Table 1). More detailed information about the participants' characteristics can be found in Table 1.

Table 1

Demographics including Frequencies or Means of Gender, Age Groups, BMI, ED pathology, Self-Criticism, and Mental Wellbeing

Characteristic/ Scale	Total <i>N</i> = 981	ED Type				$\Delta\chi^2$
		AN <i>N</i> = 295 (30.1%)	BN <i>N</i> = 220 (22.4%)	BED <i>N</i> = 89 (9.1%)	OSFED <i>N</i> = 377 (38.4%)	
	Freq. (%)	Freq. (%)	Freq. (%)	Freq. (%)	Freq. (%)	
Gender						χ^2 (3) = 10.30*
Female	954 (97.2%)	291 (98.6%)	218 (99.1%)	85 (95.5%)	360 (95.5%)	
Age Group						χ^2 (21) = 98.23*
17-30	707 (72.1%)	245 (83%)	158 (71.8%)	44 (49.4%)	260 (69%)	
31-50	242 (24.7%)	44 (14.9%)	59 (26.8%)	36 (40.4%)	103 (27.4%)	
51-61+	32 (3.2%)	6 (2.1%)	3 (1.4%)	9 (10.1%)	14 (3.7%)	
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	F
BMI	22.69 (7.63)	16.62 (1.58)	23.77 (4.73)	35.43 (8.78)	23.82 (7.04)	-
EDE-Q	3.91 (1.16)	3.80 (1.17)	4.15 (1.08)	3.66 (1.03)	3.89 (1.21)	F (3) = 5.51***
FSCRS						
IS	25.11 (6.42)	25.26 (6.42)	26.12 (6.12)	25.11 (6.57)	24.41 (6.51)	F (3) = 3.41*
HS	8.17 (5.00)	8.58 (5.23)	8.60 (4.90)	8.12 (4.96)	7.60 (4.83)	F (3) = 2.90*
RS	12.75 (5.56)	12.09 (5.38)	12.69 (5.24)	12.75 (6.06)	13.31 (5.72)	F (3) = 2.67*
MHC-SF	2.41 (.88)	2.34 (.92)	2.25 (.88)	2.50 (.89)	2.53 (.84)	F (3) = 5.85***
Psychological	2.49 (1.00)	2.38 (1.01)	2.32 (.98)	2.64 (.98)	2.64 (.98)	F (3) = 6.94**
Emotional	2.56 (1.06)	2.41 (1.13)	2.48 (1.05)	2.64 (1.12)	2.68 (.99)	F (3) = 4.20**
Social	2.22 (.98)	2.25 (.98)	2.02 (.98)	2.26 (1.05)	2.31 (.95)	F (3) = 4.25**

Note. * $p < .05$; ** $p < .01$; *** $p < .001$; ED = eating disorder; AN = Anorexia Nervosa; BN = Bulimia Nervosa; BED = Binge Eating Disorder; OSFED = Other Specified Feeding or Eating Disorder; BMI = Body Mass Index; EDE-Q = Eating Disorder Examination Questionnaire; FSCRS = The Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale; MHC-SF = Mental Health Continuum – Short Form; IS = Inadequate Self; HS = Hated Self; RS = Reassured Self; Freq. = Frequency; M = Mean; SD = Standard Deviation; $\Delta\chi^2$ = Chi-Square Difference Test.

Hypothesis 1: Confirmatory Factor Analysis

Fit indices imply a good fit of the three-factor solution. Most measures of fit indices met the criteria (Table 3). Yet, the Chi-Square test of model fit of the three-factor solution was significant. Regarding the two-factor structure, although the criteria were not directly met, the CFA demonstrated reasonable fit of the model (Table 2). Again, the Chi-Square test of model fit was significant as well. When comparing the models, both, the fit indices (i.e., CFI, TLI, RMSEA, SRMR) and the Chi-Squared Difference Test ($\Delta\chi^2 = 400.02$; $\Delta df = 2$, $p < .001$) indicated a better fit for the three-factor solution. Consequently, Hypothesis 1 can be confirmed, and further multigroup confirmatory factor analysis will be conducted using the three-factor solution. Table 2 offers an overview of the two model fit indices as well as difference test.

The standardised factor loadings of the three-factor solution, including Factor 1.1 (Inadequate Self), Factor 1.2 (Hated Self), and Factor 1.3 (Reassured Self), can be found in Table 3. For the two-factor solution, including Factor 2.1 (Self-Criticism) and Factor 2.2 (Self-Reassurance), standardised factor loadings can be found in Appendix D. For both factor solutions only two items (Item 4, Item 19) were found to be below the afore established threshold of .4 and thus inherit unsatisfactory loadings.

Table 2

CFA fit indices of the three- and -two factor solution, and model comparison

Model	χ^2	CFI	TLI	RMSEA	SRMR	$\Delta \chi^2$
Three-Factor	$\chi^2 (206) = 1018.41^{***}$.91	.90	.06	.05	-
Two-Factor	$\chi^2 (208) = 1418.43^{***}$.87	.85	.06	.08	-
Three-factors vs. Two-factors	-	-	-	-	-	$\Delta \chi^2 = 400.02$; $\Delta df = 2^{***}$

Note. $^{***}p < .001$; χ^2 = Chi-Square Test of Model Fit; $\Delta\chi^2$ = Chi-Square Difference Test; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; $\Delta\chi^2$ = Chi-Square Difference Test.

Table 3*Standardised factor loadings for the three-factor solution*

Item	Factor Loading		
	Factor 1.1: IS	Factor 1.2: HS	Factor 1.3: RS
1. I am easily disappointed with myself	.79		
2. There is a part of me that puts me down	.83		
4. I find it difficult to control my anger and frustration at myself	.37		
6. There is a part of me that feels I am not good enough	.72		
7. I feel beaten down by own self-critical thoughts	.71		
14. I remember and dwell on my failings	.57		
17. I can't accept failures and setbacks without feeling inadequate	.50		
18. I think I deserve my self-criticism	.63		
20. There is a part of me that wants to get rid of the bits I don't like	.45		
9. I have become so angry with myself that I want to hurt or injure myself		.60	
10. I have a sense of disgust with myself		.76	
12. I stop caring about myself		.73	
15. I call myself names		.63	
22. I do not like being me		.76	
3. I am able to remind myself of positive things about myself			.67
5. I find it easy to forgive myself			.63
8. I still like being me			.82
11. I can feel lovable and acceptable			.67
13. I find it easy to like myself			.72
16. I am gentle and supportive with myself			.65
19. I am able to care and look after myself			.31
21. I encourage myself for the future			.52

Note. IS = Inadequate Self; HS = Hated Self; RS = Reassured Self.

Hypothesis 2: Measurement Invariance with Multigroup Confirmatory Factor Analysis

For MGCFA, first individual CFAs per group were conducted. Afterwards, MGCFA was conducted on three levels namely, configural, metric and scalar level.

CFA per ED

After examining model fit in each group, all model fits were slightly below the determined cut-off points, except the BN group (see Table 4). Although this is still in a reasonable range, modification indices were calculated. Several error correlations were allowed after which the model showed good fit in all groups. Items that were correlated are also presented in Table 4.

Configural Model Test

Generally, measures of fit indices were slightly below the afore determined cut-off points (Table 4). As they are still in the range of not being considered substantially different, these values can be considered as adequate fit. Nonetheless, modification indices were calculated. After allowing error correlation between Item 17 and Item 18, the model showed an improved fit (CFI = .91; TLI = .90; RMSEA = .07; SRMR = .06) and met the criteria, although the Chi-Square test of model fit was still significant ($\chi^2(820) = 1694.46^{***}$). Notably, Item 18 showed to be correlating with Factor 2 (HS) as well, which it is not a part of initially. For this first step, measurement invariance, in terms of similar loading patterns in all ED groups, can be considered.

Metric Model Test

Similarly to the prior model test, fit indices indicated to be just meeting the cut-off points and are thus still in an acceptable range (CFI = .91; TLI = .90; RMSEA = .06; SRMR = .06). Although, the Chi-Square difference test between the configural and metric model was significant ($\Delta\chi^2=82.23$; $\Delta df = 57$, $p < .001$), the difference did not exceed the before established cut-off points and thus both models fit equally well (Table 4). Consequently, factor loadings can be assumed to be equally fitted between groups, indicating weak measurement invariance.

Scalar Model Test

For the final model, the initial model fit indices decreased (i.e., CFI) or increased (i.e., SRMR) slightly but can still be considered acceptable (CFI = .90; TLI = .90; RMSEA = .06; SRMR = .07). Finally, the Chi-Square difference test was non-significant ($\Delta\chi^2= 41.36$; $\Delta df = 53$, $p > .001$), implying that there is no substantial difference between the scalar and metric model. Also, the afore determined cut-off points for substantial model differences are not exceeded (Table 4). Thereby it is indicated that there is measurement invariance across the different groups, which confirms hypothesis two. Table 4 offers an overview of all MGCFAs model fit indices. As Item 18 (part of IS) showed to be problematic (i.e., covariance with Item 17 as well as factor 2 (HS)), it was decided to test whether freeing its factor loading might

show improvements of model fit. However, no improvements could be remarked (see Table A2 in Appendix E).

Table 4

MGCFA fit indices per ED and model level and comparison between fit indices

MGCFA	χ^2	CFI	TLI	RMSEA	SRMR	Mod. Ind.
ED 3-Factor						
AN	$\chi^2 (206) = 496.55^{***}$.90	.88	.07	.06	17, 18
AN2	$\chi^2 (205) = 470.41^{***}$.91	.89	.07	.06	
BN	$\chi^2 (206) = 345.05^{***}$.93	.92	.06	.06	-
BED	$\chi^2 (206) = 302.60^{***}$.89	.88	.07	.07	6, 7
BED2	$\chi^2 (205) = 286.80^{***}$.91	.90	.07	.07	
OSFED	$\chi^2 (206) = 623.56^{***}$.89	.87	.07	.06	15, 11
OSFED2	$\chi^2 (205) = 600.48^{***}$.90	.80	.07	.06	
Total sample						
Configural	$\chi^2 (824) = 3862.10^{***}$.90	.89	.07	.06	17, 18
Configural2	$\chi^2 (820) = 1694.46^{***}$.91	.90	.07	.06	
Metric	$\chi^2 (877) = 1776.86^{***}$.91	.90	.06	.06	-
Scalar	$\chi^2 (934) = 1891.35^{***}$.90	.90	.06	.07	-
Model	$\Delta \chi^2$	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR	
Comparison						
Configural vs. Metric	$\Delta \chi^2 = 82.23; \Delta df = 57^{***}$.003	-.004	.001	-.01	-
Metric vs. Scalar	$\Delta \chi^2 = 41.36; \Delta df = 53$.006	0	0	-.002	-

Note. $***p < .001$; MGCFA = Multigroup confirmatory factor analysis; ED = eating disorder; AN = Anorexia Nervosa; BN = Bulimia Nervosa; BED = Binge Eating Disorder; OSFED = Other Specified Feeding or Eating Disorder; χ^2 = Chi-Square Test of Model Fit; $\Delta \chi^2$ = Chi-Square Difference Test; Mod.Ind. = Modification Indices, Items of the FSCRS with the strongest covariance; AN2, BED2, OSFED2, & Configural2 = models after error correlations were allowed; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual; Δ CFI, Δ TLI, Δ RMSEA, Δ SRMR = difference of compared scores.

Total	IS	25.11 (6.42)	-.47**	-.45**	-.41**	-.37**	8.33%
Sample	HS	8.17 (5.0)	-.57**	-.55**	-.53**	-.41**	66.67%
	RS	12.75 (5.56)	.71**	.72**	.62**	.52**	100%
AN	IS	25.26 (6.42)	-.49**	-.46**	-.40**	-.43**	0%
	HS	8.58 (5.23)	-.60**	-.56**	-.55**	-.51**	100%
	RS	12.09 (5.38)	.75**	.75**	.66**	.59**	100%
AN %							66.67
BN	IS	26.12 (6.12)	-.39**	-.39**	-.35**	-.28**	0%
	HS	8.06 (4.90)	-.53**	-.54**	-.45**	-.40**	33.33%
	RS	12.69 (5.24)	.69**	.71**	.59**	.51**	100%
BN %							44.44%
BED	IS	25.11 (6.57)	-.51**	-.46**	-.51**	-.36**	33.33%
	HS	8.12 (4.96)	-.59**	-.59**	-.56**	-.37**	66.67%
	RS	12.75 (6.06)	.72**	.73**	.59**	.52**	100%
BED %							66.67%
OSFED	IS	24.42 (6.51)	-.48**	-.46**	-.42**	-.37**	0%
	HS	7.60 (4.83)	-.54**	-.52**	-.54**	-.35**	66.67%
	RS	13.31 (5.72)	.70**	.70**	.61**	.45**	66.67%
OSFED %							44.44%
Total %							55.56

Note. ** $p < .01$; % = percentage of correlations meeting the cut-off point of .5; AN = Anorexia Nervosa; BN = Bulimia Nervosa; BED = Binge Eating Disorder; OSFED = Other Specified Feeding or Eating Disorder; FSCRS = The Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale; MHC-SF = Mental Health Continuum – Short Form; M = Mean; SD = Standard Deviation.

Hypothesis 4: Convergent Validity between Groups

Even though convergent validity was not found, all correlations were significant and most showed to be adequate to strong. Hence, correlations between groups were compared, nonetheless. General Linear Model analysis revealed no significant results, indicating that there is no correlational difference between the different ED types (see Table 6). Hence, the final hypothesis can be accepted.

Table 6

GLM of FSCRS and MHC-SF correlations

FSCRS Subscales	MHC-SF (Sub)scale	df	F	p
IS	Total	3	.74	.53
	Psychological	3	.35	.79
	Emotional	3	.75	.52
	Social	3	1.00	.39
HS	Total	3	.46	.71
	Psychological	3	.23	.88
	Emotional	3	1.00	.38
	Social	3	1.75	.16
RS	Total	3	.36	.78
	Psychological	3	.25	.86
	Emotional	3	.43	.74
	Social	3	.75	.52

Note. GLM = General Linear Model analysis; IS = Inadequate Self; HS = Hated Self; RS = Reassured Self; FSCRS = The Forms of Self-Criticizing/ - Attacking and Self-Reassuring Scale; MHC-SF = Mental Health Continuum – Short Form.

Discussion

This study aimed at adding to existing knowledge by further examining the psychometric properties of the FSCRS. It is the first to examine measurement invariance and convergent validity with the MHC-SF between different ED types (i.e., AN, BN, BED, OSFED). Overall, results were promising as measurement invariance could be identified, whereby the initially intended three-factor structure showed a better fit than solely two factors. Added to that, although significant, and partially strong, correlations could be identified, this study was unable to identify sufficient convergent validity between the self-

criticism subscales and mental wellbeing. For self-reassurance however, convergent validity was found. Further analysis showed no differences in correlations between groups.

Theoretical Reflection

According to the obtained results, the first, second and fourth hypothesis can be confirmed. In the following, the results will be further explained and reflected upon with relevant literature.

Hypothesis 1

Starting with the first hypothesis that a three-factor solution (IS, HS, and RS) will provide a better fit compared to a two-factor solution (self-criticism and RS) in this study can be confirmed. Consequently, in this context, self-criticism did consist of two components as initially intended by Gilbert and colleagues (2004), namely Hated Self and Inadequate Self. Therefore, fit indices of the three-factor solution generally showed a good fit. Indicating that (at least in a clinical sample) self-criticism, as measured using this scale, represents two distinctive patterns of thoughts/beliefs in people with EDs. Thereby distinguishing reliably between criticism of one's performance and one's own self. On the other hand, results of the two-factor solution were not able to meet the afore established criteria which makes it less suitable to represent the factor structure of the FSCRS in this study.

This is also in line with prior research, for instance conducted by Baião et al. (2015) or Castilho et al. (2015). Notably, fit indices of the two-factor solution were arguably close to identify an acceptable fit which might be an indicator that items per scale are highly correlated across factors and hence not that easily assignable to one component of self-criticism. In connection with that, Halamová and colleagues (2018) found the two-factor solution to be best suited. However, they emphasised that their findings still need to be further empirically supported and that it needed to be differentiated between clinical- and non-clinical populations. Thereby, they suggested using three-factors for the clinical group and a two-factor structure for non-clinical populations (Halamová et al., 2018). As this study solely concerns a clinical sample, this does also support the findings of a superior three-factor structure fit.

Hypothesis 2

Secondly, it was hypothesised that self-criticism is conceptualised similarly among the different ED types. There are no studies that examined whether there are differences in the conceptualisation of self-criticism in EDs to this point. Still, it was expected that MGCFA would be able to identify measurement invariance on three levels (i.e., configural, metric, and scalar). Thereby, it was tested whether the FSCRS inherits the same factorial structure in four

different ED types. As results have shown, measurement invariance could be confirmed on all three levels. Hence, the same three-factor structure applied in all four groups was holding up to more and more strict model requirements. This indicates that across the different types of EDs, participants of the study did not significantly differ in which items they endorsed or did not endorse. Thereby suggesting that self-criticism is conceptualised similarly in people with EDs regardless of the specific type and that the underlying latent factors are the same across groups (Bellur et al., 2021).

Nonetheless, even though these findings are promising, it needs to be considered that they were just able to meet the afore determined criteria (Pirralha, 2020). This, in addition to being the first examination in this regard, gives reason to replicate MGCFA in this specific population. Arguably, the subsamples had significantly different FSCRS (sub)scale scores, which might also have an effect on the results, although differences are seemingly small. However, in prior research, MGCFA of the FSCRS was only conducted in mixed-clinical and non-clinical samples (e.g., Bellur et al., 2021; Castilho et al., 2015). These results were able to demonstrate measurement invariance, meaning that the FSCRS identified characteristics of self-criticism and -reassurance on the same factorial structure in clinically diverse samples (Bellur, et al., 2021; Baião et al., 2015). Following that, it is unlikely that the sub-samples characteristics had such a meaningful effect on the results. Notably, there is no direct literature to compare these findings to. Added to that, the group size of the different ED types might not have been ideal in terms of BED patients making up a low portion of the whole sample. Thereby, this group was below the preferred size of 100 (Memon et al., 2020). Still, the results for this group were highly comparable to the other diagnoses.

Hypothesis 3 and 4

It was expected that convergent validity of the FSCRS and mental wellbeing scale could be demonstrated. Thereby self-criticism and mental wellbeing were thought to show negative strong correlations, while self-reassurance possesses a strong positive correlation with mental wellbeing. This hypothesis could not be fully supported, although all correlations were significant and mostly adequate to strong. For the total sample only the combined RS correlations were able to exceed the 75% mark, which allows considering convergent validity (Abma et al., 2016). In sum, just 55.56% of all correlations calculated were able to exceed the afore established cut-off of .5, to be considered a strong enough correlation.

Especially values of the IS-subscale did not exceed the cut-off point of .5. Hence, convergent validity cannot be considered. However, the correlation with total mental wellbeing is arguably close to it ($r = -.47, p < .001$). Even though not meeting 75% either, HS

still performed better in terms of higher correlations, whereby 66.67% were higher than .5. This is mostly due to HS meeting the criteria with 100% in the AN group. Hereby, a possible explanation may be given by the assumption of Gilbert and colleagues (2004) that the HS might represent a more pathological domain, which can be associated with self-harm or depressive mood. Considering that the AN group's wellbeing scores were, compared to the other groups, significantly lower (except BN), there might be a link from HS to lower mental wellbeing and hence stronger negative correlations. However, there are no direct data on self-harming tendencies or depressive mood for this sample. Thus, this reasoning needs to be treated with caution.

Nonetheless, this does not necessarily contradict already conducted research. For instance, Sommers-Spijkerman and colleagues (2018) did confirm convergent validity in a non-clinical sample but did not establish a certain cut-off point. When comparing values directly, Sommers-Spijkerman et al.'s (2018) correlations are weak to moderate. Yet, the sample of their study already indicated lowered FSCRS scores, and self-criticism levels of this study's sample was almost twice as high. Nonetheless, wellbeing scores appear to be similar although participants of Sommers-Spijkerman et al. (2018) had slightly higher scores. Moreover, Becker (2018) showed similar results to this study, but also considering convergent validity when correlations showed to be moderate and significant. Arguably, a quite high cut-off point to consider convergent validity was applied in the present study. Following this, if moderate correlations were also considered as convergent validity as well, it is to be expected that convergent validity would have been determined.

Markedly, RS had the strongest correlations with all wellbeing subscales in this and the reference studies (Sommers-Spijkerman et al., 2018; Becker, 2018). This matches with the results of the study by Duarte and colleagues (2017), in which self-reassurance could be linked with higher mental wellbeing of obese individuals that were attempting to manage their weight. More specifically, the researchers found that the relationship between self-reassurance and wellbeing was stronger than the negative association of self-criticism and wellbeing (Duarte et al., 2017).

Finally, there were no differences regarding correlations between self-criticism and mental wellbeing between ED types expected. Generally, this hypothesis can be confirmed as no significant differences between the groups regarding correlations were identified. As for Hypothesis 2, this is the first study to compare the convergent validity of the FSCRS and MHC-SF between different ED types. As for validation purposes, this result also implies that the FSCRS can be applied across different ED types with varying disorder presentations.

Strengths and Limitations

All in all, several strengths of this study can be named. First off, as this study only measured at one point in time, there were relatively few participants that needed to be excluded due to missing data, which enabled the researchers to collect a number of various measures of a considerable big sample size. Meaning that it can be expected that this study inherits a certain degree of significance (Kyriazos, 2018). Added to that, there was a relative balance of participating ED patient types, which should give a good insight into the comparison of these types. Moreover, the used questionnaires all possess good reliability generally and thus the obtained results can be accredited (George & Mallery, 2003). A last strength of this study is that validity was assessed in a solely clinical sample, in a patient group in which their disorders are often collectively regarded to. Furthermore, analyses were conducted for the whole sample as well as the diagnosis groups individually.

Still, as with any research, this current study also contains some limitations. First, there is generally a very uneven distribution of genders, which makes the sample quite homogenous. Females inherit the vast majority, which makes the results mostly representative for female ED patients. Even though females have higher prevalent rates of EDs compared to males (Galmiche et al., 2019), males in this study are still comparatively few. Especially, in the BN group there is a large deviation as males make up < 1% of the sub-sample, which makes it difficult to generalise the findings for this group specifically. However, an uneven distribution of genders is very common in ED research and as such the results as a whole can be accredited (e.g., Kelly & Carter, 2013; Kelly & Tasca, 2016). Another limitation of this study was the comparatively low number of participating BED patients (9%). As a result, MGCFA for the BED group might not be as meaningful as for the other ED groups as the number of BED patients was below the preferred group number of 100 to perform MGCFA (Memon et al., 2020). Lastly, the cut-off point for convergent validity was set very high, which interfered with determining convergent validity. Arguably, a correlation of .5 was the only set, although arbitrary, cut-off point that was mentioned in the reviewed literature (Abma et al., 2016). Still, also the authors mention that it is unlikely to expect such high correlations between scales not measuring the same constructs. Consequently, a closer orientation on previous literature conducting similar analyses, such as Sommers-Spijkerman and colleagues (2018) would have been more beneficial.

Implications for Practice and Future Research

Overall, these findings suggest that when assessing self-criticism with the FSCRS, this tool can be applied to various types of ED patients. Yet, it needs to be considered that according to Memon and colleagues (2020), there were too few participants in the BED group, which usually would have been a criterion for exclusion. For this reason, it is advisable to repeat this analysis with an appropriate number of participants. Also, it should be assessed whether these findings can be replicated generally as this is the first in this context. Still, these findings are promising and the results themselves indicate that the scale holds the same significance among variable ED presentations.

Next to that, although convergent validity itself could not be established in this study for all subscales of the FSCRS, RS largely met the set cut-off point and highly correlated with mental wellbeing in all EDs. Also, for IS and HS values were still arguably strong and significant. This underlines the relationship of ED patients' self-critical attitudes and their mental wellbeing. So, as already discussed, in replicational studies, the cut-off point for convergent validity should not be set as high and rather kept on a moderate strength level. Besides, since there seemed to no difference in correlations between EDs, it can be argued that the FSCRS is equally valid for the examined ED types in this regard as well.

References

- Aardoom, J. J., Dingemans, A. E., Op't Landt, M. C. S., & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating behaviors, 13*(4), 305-309. <https://doi.org/10.1016/j.eatbeh.2012.09.002>
- Abma, I. L., Rovers, M., & van der Wees, P. J. (2016). Appraising convergent validity of patient-reported outcome measures in systematic reviews: constructing hypotheses and interpreting outcomes. *BMC Research Notes, 9*(1), 1-5. DOI 10.1186/s13104-016-2034-2
- Baião, R., Gilbert, P., McEwan, K., & Carvalho, S. (2015). Forms of self-criticising/attacking & self-reassuring scale: Psychometric properties and normative study. *Psychology and Psychotherapy: Theory, Research and Practice, 88*(4), 438-452. <https://doi.org/10.1111/papt.12049>
- Bellur, Z., Alpay, E. H., & Aydin, A. (2021). Reliability and validity of the Turkish version of forms of self-criticizing/attacking and self-reassuring scale (FSCRS) in clinical and non-clinical samples. *Current Psychology, 1-12*. <https://doi.org/10.1007/s12144-021-01408-0>
- Brennan, M. A., Emmerling, M. E., & Whelton, W. J. (2014). Emotion-focused group therapy: Addressing self-criticism in the treatment of eating disorders. *Counselling and Psychotherapy Research, 1-9*. <https://doi.org/10.1080/14733145.2014.914549>
- Castilho, P., Pinto-Gouveia, J., & Duarte, J. (2015). Exploring self-criticism: Confirmatory factor analysis of the FSCRS in clinical and nonclinical samples. *Clinical psychology & psychotherapy, 22*(2), 153-164. <https://doi.org/10.1002/cpp.1881>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal, 14*(3), 464-504. <https://doi.org/10.1080/10705510701301834>

- Duarte, C., & Pinto-Gouveia, J. (2017). Can self-reassurance buffer against the impact of bullying? Effects on body shame and disordered eating in adolescence. *Journal of child and family studies*, 26(4), 998-1006. <https://doi.org/10.1007/s10826-016-0624-1>
- Duarte, C., Ferreira, C., & Pinto-Gouveia, J. (2016). At the core of eating disorders: Overvaluation, social rank, self-criticism and shame in anorexia, bulimia and binge eating disorder. *Comprehensive psychiatry*, 66, 123-131. <https://doi.org/10.1016/j.comppsy.2016.01.00>
- Dunkley, D. M., Masheb, R. M., & Grilo, C. M. (2010). Childhood maltreatment, depressive symptoms, and body dissatisfaction in patients with binge eating disorder: The mediating role of self-criticism. *International Journal of Eating Disorders*, 43(3), 274-281. <https://doi.org/10.1002/eat.20796>
- Eid, M., Gollwitzer, M., & Schmitt, M. (2017). *Statistik und forschungsmethoden*.
- Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *International journal of eating disorders*, 16(4), 363-370. [https://doi.org/10.1002/1098-108X\(199412\)16:43.0.CO;2-%23](https://doi.org/10.1002/1098-108X(199412)16:43.0.CO;2-%23)
- Fairburn, C. G., Cooper, Z., & Shafran, R. (2003). Cognitive behavior therapy for eating disorders: A “transdiagnostic” theory and treatment. *Behaviour Research and Therapy*, 41(5), 509–528. [https://doi.org/10.1016/S0005-7967\(02\)00088-8](https://doi.org/10.1016/S0005-7967(02)00088-8)
- Fennig, S., Hadas, A., Itzhaky, L., Roe, D., Apter, A., & Shahar, G. (2008). Self-criticism is a key predictor of eating disorder dimensions among inpatient adolescent females. *International Journal of Eating Disorders*, 41(8), 762-765. <https://doi.org/10.1002/eat.20573>
- Galmiche, M., Déchelotte, P., Lambert, G., & Tavolacci, M. P. (2019). Prevalence of eating disorders over the 2000–2018 period: a systematic literature review. *The American journal of clinical nutrition*, 109(5), 1402-1413. <https://doi.org/10.1093/ajcn/nqy342>

- George, D., & Mallery, P. (2003). Reliability analysis. SPSS for Windows, step by step: a simple guide and reference. Boston: Allyn & Bacon, 222, 232.
- Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, *10*(2), 486. doi: 10.5812/ijem.3505
- Gilbert, P., Clarke, M., Hempel, S., Miles, J. N., & Irons, C. (2004). Criticizing and reassuring oneself: An exploration of forms, styles and reasons in female students. *British Journal of Clinical Psychology*, *43*(1), 31-50.
<https://doi.org/10.1348/014466504772812959>
- Halamová, J., Kanovský, M., Gilbert, P., Troop, N. A., Zuroff, D. C., Hermanto, N., ... & Kupeli, N. (2018). The factor structure of the forms of self-criticising/attacking & self-reassuring scale in thirteen distinct populations. *Journal of Psychopathology and Behavioral Assessment*, *40*(4), 736-751. <https://doi.org/10.1007/s10862-018-9686-2>
- IBM Corp. Released 2021. IBM SPSS Statistics for Macintosh, Version 28.0. Armonk, NY: IBM Corp
- Kelly, A. C., & Carter, J. C. (2013). Why self-critical patients present with more severe eating disorder pathology: The mediating role of shame. *British Journal of Clinical Psychology*, *52*(2), 148-161. <https://doi.org/10.1111/bjc.12006>
- Kelly, A. C., & Tasca, G. A. (2016). Within-persons predictors of change during eating disorders treatment: An examination of self-compassion, self-criticism, shame, and eating disorder symptoms. *International Journal of Eating Disorders*, *49*(7), 716-722. <https://doi.org/10.1002/eat.22527>
- Keyes, C. L. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Social Behavior*, 207-222. Retrieved from <http://www.jstor.org/stable/3090197>

- Kyriazos, T. A. (2018). Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(08), 2207. Doi: 10.4236/psych.2018.98126
- 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 (MHCSF). *Journal of Clinical Psychology*, 67(1), 99-110. doi: 10.1002/jclp.20741
- Löw, C. A., Schauenburg, H., & Dinger, U. (2020). Self-criticism and psychotherapy outcome: A systematic review and meta-analysis. *Clinical psychology review*, 75, 101808. <https://doi.org/10.1016/j.cpr.2019.101808>
- Memon, M. A., Ting, H., Cheah, J. H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: review and recommendations. *Journal of Applied Structural Equation Modeling*, 4(2), 1-20. DOI:10.47263/JASEM.4(2)01
- Mendes, A. L., Ferreira, C., & Trindade, I. A. (2019). The central role of self-reassurance to explain body and eating attitudes. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 24(5), 861-868. <https://doi.org/10.1007/s40519-018-0568-0>
- Munro, C., Randell, L., & Lawrie, S. M. (2017). An integrative bio-psycho-social theory of anorexia nervosa. *Clinical psychology & psychotherapy*, 24(1), 1-21. <https://doi.org/10.1002/cpp.2047>
- Navarrete, J., Herrero, R., Soler, J., Domínguez-Clavé, E., Baños, R., & Cebolla, A. (2021). Assessing self-criticism and self-reassurance: Examining psychometric properties and clinical usefulness of the Short-Form of the Forms of Self-Criticizing/Attacking & Self-Reassuring Scale (FSCRS-SF) in Spanish sample. *Plos one*, 16(5), e0252089. <https://doi.org/10.1371/journal.pone.0252089>
- Noordenbos, G., Aliakbari, N., & Campbell, R. (2014). The relationship among critical inner voices, low self-esteem, and self-criticism in eating disorders. *Eating disorders*, 22(4), 337-351. <https://doi.org/10.1080/10640266.2014.898983>

- Perkins, N. M., Ortiz, S. N., & Smith, A. R. (2020). Self-criticism longitudinally predicts nonsuicidal self-injury in eating disorders. *Eating disorders*, 28(2), 157-170.
<https://doi.org/10.1080/10640266.2019.1695450>
- Pirralha, A. (2020). Testing for measurement invariance with many groups. In *Workshop at the WAPOR 73rd Annual Conference, Virtual Conference*.
<https://www.andrepirralha.com/talk/invariance>.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://www.jstatsoft.org/v48/i02/>.
- Serpell, L., Amey, R., & Kamboj, S. K. (2020). The role of self-compassion and self-criticism in binge eating behaviour. *Appetite*, 144, 104470.
<https://doi.org/10.1016/j.appet.2019.104470>
- Sommers-Spijkerman, M., Trompetter, H., Ten Klooster, P., Schreurs, K., Gilbert, P., & Bohlmeijer, E. (2018). Development and validation of the forms of Self-Criticizing/Attacking and Self-Reassuring Scale—Short Form. *Psychological Assessment*, 30(6), 729. <https://psycnet.apa.org/doi/10.1037/pas0000514>
- Sutlive, V. H., & Ulrich, D. A. (1998). Interpreting statistical significance and meaningfulness in adapted physical activity research. *Adapted Physical Activity Quarterly*, 15(2), 103-118. <https://doi.org/10.1123/apaq.15.2.103>
- Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Measurement invariance. *Frontiers in psychology*, 6, 1064.
<https://doi.org/10.3389/fpsyg.2015.01064>
- Van Furth, E. F. (2000). Nederlandse vertaling van de EDE-Q. Leidschendam: Robert-Fleury Stichting.

Williams, B. M., & Levinson, C. A. (2022). A model of self-criticism as a transdiagnostic mechanism of eating disorder comorbidity: A review. *New Ideas in Psychology, 66*, 100949. <https://doi.org/10.1016/j.newideapsych.2022.100949>

Appendix A

The Eating Disorder Examination Questionnaire

https://www.corc.uk.net/media/1273/ede-q_questionnaire.pdf

Appendix B

Forms of Self-Criticizing/ - Attacking and Self-reassuring Scale Questions

-
1. I am easily disappointed with myself
 2. There is a part of me that puts me down
 3. I am able to remind myself of positive things about myself
 4. I find it difficult to control my anger and frustration at myself
 5. I find it easy to forgive myself
 6. There is a part of me that feels I am not good enough
 7. I feel beaten down by own self-critical thoughts
 8. I still like being me
 9. I have become so angry with myself that I want to hurt or injure myself
 10. I have a sense of disgust with myself
 11. I can feel lovable and acceptable
 12. I stop caring about myself
 13. I find it easy to like myself
 14. I remember and dwell on my failings
 15. I call myself names
 16. I am gentle and supportive with myself
 17. I can't accept failures and setbacks without feeling inadequate
 18. I think I deserve my self-criticism
 19. I am able to care and look after myself
 20. There is a part of me that wants to get rid of the bits I don't like
 21. I encourage myself for the future
 22. I do not like being me
-

Appendix C

Mental Health Continuum Short-Form Questionnaire

How often in the past month did you feel ...

1. happy?
 2. interested in life?
 3. satisfied with your life?
-

-
4. that you had something important to contribute to society?
 5. that you belonged to a community (like a social group or your neighbourhood)?
 6. that our society is a good place, or is becoming a better place, for all people?
 7. that people are basically good?
 8. that the way our society works makes sense to you?
 9. that you liked most parts of your personality?
 10. good at managing the responsibilities of your daily life?
 11. that you had warm and trusting relationships with others?
 12. that you had experiences that challenged you to grow and become a better person?
 13. confident to think or express your own ideas and opinions?
 14. that your life has a sense of direction or meaning to it?
-

Appendix D

Table with Standardised Factor Loadings for the Two-Factor Solution

Table A1

Standardised Factor Loadings Two-Factor Solution

Item	Factor Loading	
	Factor 2.1: Self-Criticism	Factor 2.2: Self-Reassurance
1. I am easily disappointed with myself	.74	
2. There is a part of me that puts me down	.77	
4. I find it difficult to control my anger and frustration at myself	.39	
6. There is a part of me that feels I am not good enough	.68	
7. I feel beaten down by own self-critical thoughts	.68	
14. I remember and dwell on my failings	.54	
17. I can't accept failures and setbacks without feeling inadequate	.51	
18. I think I deserve my self-criticism	.66	
20. There is a part of me that wants to get rid of the bits I don't like	.43	
9. I have become so angry with myself that I want to hurt or injure myself	.54	
10. I have a sense of disgust with myself	.70	

12. I stop caring about myself	.65	
15. I call myself names	.62	
22. I do not like being me	.70	
3. I am able to remind myself of positive things about myself		.68
5. I find it easy to forgive myself		.64
8. I still like being me		.81
11. I can feel lovable and acceptable		.67
13. I find it easy to like myself		.72
16. I am gentle and supportive with myself		.66
19. I am able to care and look after myself		.31
21. I encourage myself for the future		.52

Appendix E

Partial Measurement Invariance

As item 18 showed to problematic in several regards (i.e., covariance with item 17 as well as factor 2 (HS) which it initially is not a part of) and the requirements for measurement invariance were only just met, it was decided to test whether freeing the factor loading of Item 18 might show improvements of model fit. As can be taken from Table A2, fit indices are still only just meeting the criteria and hence no improvement can be remarked.

Table A2

Partial Invariance Fit Indices

Model	χ^2	CFI	TLI	RMSEA	SRMR	Mod. Ind.
Metric2	$\chi^2 (874) = 1770.80^{***}$.90	.90	.07	.07	F1 =~ 18
Scalar2	$\chi^2 (931) = 1885.60^{***}$.90	.90	.07	.07	F1 =~ 18

Note. =~ =loading of item was freed.