From Critique to Care: A New Lexicon for Understanding Self-Compassion and Self-Criticism

Joy Libosan (s2605163)

Department of Psychology, University of Twente

Dr. Marcel Pieterse & Dr. Gerben Westerhof

April 3, 2025

Word count: 10938

Table of Contents

Abstract

This thesis sought to develop a distinctive lexicon tailored for evaluating self-compassion (SC) and self-criticism (SCr) in a text-based framework. This study aimed to overcome the limits of existing measurement tools regarding the significance of these domains in psychological well-being. The newly constructed dictionary was assessed for validity and reliability using a mixed-methods approach, which included a user-defined dictionary (UDD) phase, test-retest reliability, and sensitivity analyses. The findings demonstrated that the lexicon displayed insufficient validity and reliability, as indicated by weak correlations with the Self-Compassion Scale (SCS). Moreover, SC and SCr were recognized as separate constructs instead of a singular dimension, indicating substantial differences in their representation. Qualitative results revealed participants' intricate self-perceptions that the lexicon insufficiently represented. This research highlights the necessity for context-specific terminology in assessing SC and SCr. The study enhances the comprehension of these constructs by incorporating qualitative insights into lexicon building, offering implications for therapeutic methods and directing future research efforts.

Keywords: self-compassion, self-criticism, psychological well-being, emotional assessment, lexicon development, mixed methods, construct validity, sensitivity to change, test-retest measures, qualitative analysis, mental health outcomes, linguistic analysis

From Critique to Care: A New Lexicon for Understanding Self-Compassion and Self-Criticism

In recent years, there has been increasing focus on the rising prevalence of mood disorders in the Netherlands due to their harmful impact on individuals and society. The Netherlands Mental Health Survey and Incidence Studies (NEMESIS) indicate that the prevalence of mood disorders, including depression, rose significantly from 17.4% to 26.1% between 2007–2009 and 2019–2022 (Have et al., 2023). Numerous studies indicate that Self-Criticism (SCr) may be an important contributing factor to this prevalence, as it correlates with various mental health disorders, including major depression, borderline personality disorder, eating disorders, and post-traumatic stress disorder (Löw et al., 2019; Cox et al., 2003; Fennig et al., 2008; Marshall et al., 2008). Moreover, SCr entails a damaging type of self-evaluation that involves experiencing shame and emotions of worthlessness, feeling like a failure, and sensing other people's rejection and judgment based on how well one performs in daily tasks (Potter et al., 2014).

As awareness of the adverse effects of mental health concerns has increased, so too has the focus on potential preventive and supportive factors that may mitigate these challenges. To do so, a focus has been put on Self-Compassion (SC) due to its promising results in reducing and preventing SCr (Wakelin et al., 2021). Moreover, this form of caring behaviour stems from the concept of compassion, which is conceptualised as having five dimensions that interconnect: recognising suffering, understanding the universality of suffering, emotional resonance, tolerating uncomfortable feelings and acting to alleviate suffering (Strauss et al., 2016). Although the development of caring behaviour is associated with both compassion and SC (Gilbert, 2005), SC is internally focused, whereas compassion is outwardly focused. Neff (2011) argued that SC comprises three facets: mindfulness (holding painful feelings in mindful awareness rather than over-identifying with them) versus

overidentification, self-kindness versus self-judgement and common humanity (understanding the universality of suffering) versus isolation (Neff, 2011). These components work together to foster an SC mindset when faced with mistakes made by oneself, feelings of inadequacy, or different challenging situations in life (Neff, 2011).

The distinction between mindfulness and overidentification entails maintaining a conscious and balanced awareness of painful emotions associated with one's suffering, thereby recognising those sufferings. However, if an individual opposes or fights these unpleasant emotions, they may become absorbed by their suffering and feel incapable of (emotionally) caring for themselves. One thus may become excessively identified with negative feelings and thoughts (Finlay-Jones et al., 2023). Further, being loving, kind, and understanding toward oneself and actively calming and comforting oneself during difficult circumstances are all components of cultivating self-kindness (Neff, 2011). This reaction contrasts with a self-critical mindset, which holds one responsible for one's own shortcomings or inadequate ability to handle obstacles in life (Neff, 2011). Lastly, common humanity versus isolation entails acknowledging one's sense of connection to the collective human experience of imperfection, cultivating a profound sense of belonging rather than feeling isolated in times of hardship (Neff, 2011).

As previously mentioned, there is seemingly an important relationship between SC and SCr, and Neff (2003) argues that the two concepts can be considered opposites of the same construct, referred to as a bipolar construct (Neff, 2003). Neff also strengthens her claim by demonstrating that all factors of her Self-Compassion Scale (SCS), load on one global factor, along with the six subfactors of the scale that represent both poles. Moreover, there is substantial data supporting the reliability and validity of the SCS (Neff, 2015). Research indicates that the internal reliability of the SCS remains consistently high across diverse populations, implying acceptable inter-correlation among all SCS items (Neff and Pommier

2013; Werner et al. 2012). The extensive corpus of studies demonstrates that scores on the SCS forecast well-being signify high predictive validity (Neff, 2015).

Text Analysis and LIWC

While a number of structured instruments, like Neff's (2011) Self-Compassion Scale (SCS), are useful for directly measuring SC and SCr through self-report, analysing real-world narratives may be a complementary but useful approach because it can offer important insights into the lived experience of psychological constructs like SC (Ponizovskiy et al., 2020; Ji & Raney, 2020). In other words, it can cast light on complex mental states or emotional experiences that one expresses concerning one's perception of oneself and one's means of coping.

Furthermore, a substantial body of research has been gathered by social scientists that emphasise the psychological significance of the words people use to express meaning and emotions, as it may provide information on personal beliefs, personality, ways of thinking, their emotions as well as lived experiences (Boyd et al., 2022). Moreover, in a qualitative study by Sevilla-Liu (2022), where she examined SC in students' life stories through narratives, she discovered that SC was not only found on the level of reported behaviours or thoughts but integrated into a person's self-perception and narrative identity. Thus, focusing on the narratives of people might provide insightful knowledge on how people express SC and SCr in written language and what words or expressions are mostly associated with these concepts.

Consequently, text-mining approaches have proven to be quite insightful when applied to psychological phenomena. To elaborate, according to Gorvankolla (2017), text mining is the process of discovering new, unexplored information by utilising computers to automatically gather information from various written resources. In other words, it pertains to the procedure of deriving meaningful patterns or knowledge from textual materials.

Moreover, the linguistic, statistical, and machine-learning methods necessary for analysing textual data for more data analysis, research, and inquiry are all included in text mining (Hashimi et al., 2015). According to the literature and applications now accessible, text mining is widely employed in a variety of fields, including theme tracking, visualisation of information, information retrieval, affect in text, etc (Hashimi et al., 2015).

One text-mining tool for counting psychological terms associated with affect, cognition, and biological processes is the Linguistic Inquiry and Word Count (LIWC) instrument (Boyd et al., 2022). Moreover, it is a straightforward text analysis tool that counts words in psychologically significant categories. Empirical studies utilising LIWC illustrate its capacity to identify meaning in an extensive range of experimental instances, including demonstrating attentional focus, affect, social interactions, cognitive styles, and personality variations (Tausczik & Pennebaker, 2009). The fundamental components of LIWC-22 are software and a "dictionary," which is a map that links significant psychological theories with words, sentences, and other linguistic constructs (Boyd et al., 2022). With the inclusion of a dictionary workbench, Users can use it to evaluate texts using dictionaries related to their own interests (Boyd et al., 2022). Stated differently, researchers possessing an approved dictionary extension are free to employ LIWC for any analytical purpose. Thus, this method allows for measuring SC in a novel way, as there are no dictionaries yet for the constructs of SC and SCr, providing new insights.

Moreover, LIWC's contributions to establishing criteria and incremental validity in text-based evaluations signify a substantial advancement in the discipline. Through comparing linguistic outputs with established psychological assessments, LIWC has shown that language usage can reliably predict personality traits, enhancing conventional self-report techniques (Tausczik & Pennebaker, 2010). In contrast to earlier approved dictionaries that concentrated on specific emotional words, LIWC's extensive collection incorporates a wide range of

linguistic patterns. This progression promotes the development of customised dictionaries that may encapsulate the subtleties associated with concepts like SC and SCr.

Next, creating a dictionary using the User-Defined Dictionaries (UDD) method is a straightforward method to investigate the relationship between language and speaker characteristics. Because the categories of interest are represented by single words and /or strings of multiple words, a computer program can find relevant content automatically throughout large corpora of texts (Iliev et al., 2014b). As a result, collections of terms related to specific themes, such as SC and SCr, can be established with theoretical adaptability and ease (Iliev et al., 2014b).

Nevertheless, The UDD method needs to be handled with caution as, when presented in its most basic form (as counts of words or strings of words that appear in specific user-defined categories), it overlooks the context in which words are found (Iliev et al., 2014b). To address this, human verification of the data and the application of machine-learning techniques, which focuses on developing algorithms that can replicate human cognitive capacities from a learning standpoint, might be beneficial in countering text-mining issues such as overlooking the context or missing nuances (Iliev et al., 2014b; Žižka et al., 2019).

Appropriate Methodology & Establishing Psychometric Quality

As aforementioned, The LIWC instrument allows users to asses texts with dictionaries relevant to their interests (Pennebaker et al., 2007). In order to assess the psychometric quality of instruments developed using dictionary methods, like the UDD approach, a particular methodology that complies with psychometric standards must be followed. This entails methodically gathering and verifying the dictionary using a predetermined set of steps established by Pennebaker et al. (2007), and are swiftly explained below.

The first step is ‘word collection’. Here, researchers collect words or synonyms from multiple sources such as Thesaurus, standard dictionaries or, for example, social media posts

in search of words that lay people use within the context of the chosen dictionary term. After a lexicon has been created, step two can be executed.

This step is called the ‘judge Rating phase’. Here, the lexicon words are rated by three to four ‘experts’ (experts in the field of the chosen dictionary term) for the goodness of fit for each word. Pennebaker et al. (2007) propose that for words to be included, all judges must gain consensus on their inclusion.

The third step is called ‘Base Rate Analyses’ and involves analysing whether words from the created dictionary are used often enough across various corpora and online sources. For this analysis, they use the Test Kitchen corpus, which contains 15.000 texts from diverse (online) sources. Words that do not occur at least once are suggested to be removed from the dictionary.

The fourth step is the ‘Candidate Word List Generation’, where they propose to identify the 5,000 words that were most frequently used in the Test Kitchen Corpus. Further, words that do not yet exist in the LIWC dictionary are presumably fit for inclusion in the new one. After this, correlations may be computed to test whether the words show an acceptable fit to the chosen dictionary term.

Another crucial step for the validation process is Step 5: ‘The Psychometric Evaluation’, where internal consistency statistics are computed for each word. Subsequently, items proven to have an unfavourable effect on internal consistency should be noted for possible omission from the dictionary's definitive version, ensuring the instrument's reliability.

Lastly, step 6 is called the ‘Refinement Phase’, where they suggest repeating the previous steps to become aware of possible mistakes (Pennebaker et al., 2007).

Aim & Hypotheses

Because SC is still lacking in its comprehensive lexicon, this thesis will aim to bridge this gap by creating and validating a dictionary for SC and SCr that will serve as a new

measurement instrument. However, prior research has demonstrated the intricacy of SC and SCr as emotional variables necessitating sophisticated measurement (Neff, 2011). This study aims to refine and develop a precise lexicon for evaluating SC and SCr, drawing on current literature and expert opinion due to the intricacy of accurately expressing these constructs and, therefore, measuring content validity. For this reason, the following research question was proposed: In what way do expert perspectives and current literature contribute to the formulation of a valid and reliable lexicon intended to assess SC and SCr in participants?

After this process, another aim is to validate the dictionary using the qualitative dataset from a study that tests the effectiveness of Mindfulness-Based Compassionate Living (MBCL) (Schuling et al., 2016). This dataset includes pre- and post-interviews of participants who underwent the MBCL intervention. Thus, this data could contain relevant information regarding an individual's increase or decrease in SC and SCr, possibly allowing for testing and validating the created dictionary. This allows for assessing 'sensitivity to Change', which entails the Identification of an effect that entails a shift, however slight an outcome resulting from intervention or developmental influences (Fok & Henry, 2015). In other words, it can detect meaningful changes over time, providing insightful knowledge.

Even though we consider SC and SCr to be two extremes of a continuum, we posed the following research question: Are SC and SCr best represented on a unidimensional or bidimensional scale? Thus, we will calculate separate -unipolar- scores for both, allowing us to test whether both concepts act as a single, unidimensional construct or two independent dimensions. Subsequently, by using their data, the following hypotheses will be tested:

H1: The mean frequency of lexicon words related to SCr is higher before the MBCL intervention than after.

H2: The mean frequency of lexicon words related to SC is higher after the MBCL intervention than before.

Additionally, the dataset by Schuling et al. (2016) contains self-reported measures of the participants on the SCS by Neff (2011). Despite the correlations between LIWC affect or emotion categories derived from written texts and individuals' self-reported affective states typically ranging from .05 to .40, with an average of approximately .15 to .20, indicating a generally weak relationship (Boyd et al., 2022), we contend that it is worthwhile, for testing the construct validity of the lexicon, to examine the correlations between participants' self-reports on the SCS and their lexicon scores on SC and SCr in order to evaluate whether the lexicon accurately reflects SC and SCr. Furthermore, assessing whether the lexicon exhibits stability within the measurements over time is important since it is essential for validating psychological instruments (Streiner & Kottner, 2014). For this reason, the following hypotheses have been stated:

H3: A significant positive correlation will be found between the scores of the SCS and the scores of the lexicon measuring SC after the MBCL intervention.

H4: A significant negative correlation will be found between the scores of the SCS and the scores of the lexicon measuring SCr after the MBCL intervention

H5: A significant positive correlation will be observed between the test-retest of the lexicon evaluating SC when measured pre and post-MBCL intervention.

H6: A significant positive correlation will be observed between the test-retest of the lexicon measuring SCr when measured at pre and post-MBCL intervention.

Lastly, in order to gain a deeper understanding of the constructs of SC and SCr, a qualitative analysis will be performed based on the following research question: In what way do participants' narratives in qualitative interviews illustrate their experiences of SC and SCr prior to and after to the MBCL intervention?

The validation process is further elaborated on in the section: 'Step Five: Psychometric Evaluation', and Information regarding the study by Schuling et al. (2016) will be elaborated on in the section below.

Methods

First, we describe the research and data used to create our corpora. Then, we describe the development and procedures step-by-step based on the LIWC.

Paper and Data Used

The paper of which the data used for this thesis is "A parallel-group, randomised controlled trial into the effectiveness of Mindfulness-Based Compassionate Living (MBCL) compared to treatment-as-usual in recurrent depression: Trial design and protocol" (Schuling et al., 2016). They aimed to ascertain whether patients with recurrent depression who had previously taken part in Mindfulness-Based Cognitive Therapy (MBCT) could benefit from Mindfulness-Based Compassionate Living (MBCL) in terms of lowering (residual) depressive symptoms and cultivating SC (Schuling et al., 2016). Additionally, the study included a qualitative part whereby a subsample (n = 24) was invited to participate in one-on-one in-depth interviews aimed at elucidating the nuances of participants' experiences before and after the intervention to provide additional information on the effectiveness of the MBCL intervention (Schuling et al., 2016). Additionally, the specified interview questions are not provided. However, they include evaluations of the MBCL's supplementary benefits

compared to Mindfulness-Based Cognitive Therapy (MBCT), participants' experiences with the MBCL, and their insights. Lastly, regarding the demographics of the qualitative sample, the study by Schuling et al. (2016) aimed for as much diversity as possible in terms of age, gender, and current state of depressive disorder, including both non-completers as well as completers of the intervention.

With this information from this subset of individuals who underwent this treatment, we hoped to test the construct validity of the created dictionary for the terms SC and SCr. Also, since there is information on the pre-and post-sessions per individual, the first two hypotheses stated earlier can be tested as well using this data. Thus, we believed this data could provide us with insightful knowledge regarding the concepts of SC and SCr, so the choice was made to use this data.

Procedure

As aforementioned, The LIWC instrument allows users to asses texts with dictionaries relevant to their interests (Pennebaker et al., 2007). They introduced six steps in their methods to create a dictionary. For this thesis, steps one, two and five of the LIWC have been used and are explained in the sections below. However, they were slightly altered to make it more feasible. Additionally, a qualitative study was conducted in which one subject's interview was carefully analysed to attain a comprehensive grasp of the constructs SC and SCr, as well as the contexts in which the terms were employed, thereby clarifying the participants' narrative.

Step One: 'Word Collection'

The first step that the LIWC describes is 'Word Collection' (Pennebaker et al., 2007). Within this step, it is necessary first to collect all possible words that align with the concepts of SC and SCr. We considered the definition of SC by Neff (2011). She defines SC as having three aspects: self-kindness, mindfulness, and common humanity, each of which has a

definition, as mentioned in the introduction. Thus, synonyms were also collected independently for each of the Neff-defined aspects. Additionally, the synonyms for the lexicon were gathered via Thesaurus, browsing online platforms, such as Reddit of specific forums whereby topics related to SC and SCr were seemingly relevant, academic literature, English and Dutch dictionaries, as well as brainstorming with friends and family who are inexperienced in the field of psychology and its jargon, to possibly grasp an understanding of laypeople language for the concepts of SC and SCr.

The first concept of the lexicon was created in English due to the finding that important synonyms based on academic literature, for instance, words gained from items of the Self-compassion Scale by Neff (2011), were mainly presented in English. For this thesis, the words generated for the lexicon derived from literature had to be translated into Dutch since the dictionary was to be tested on Dutch interview data by the paper of Schuling et al. (2016). The initial list of words created for SC and SCr, as well as the translation process, can be viewed in Appendix A. However, a preview of this process is presented in Table 1.

Table 1.

List of synonyms for self-compassion and their translations (self-kindness, mindfulness, common humanity)

Synonyms	Translation	Source
Compassion	Mededogen	Dutch psy. articles
Kindness	Vriendelijkheid	Van Dale
Self-kindness	Zelfvriendelijkheid	Dutch linguistic rules
Understanding	Begrip	Thesaurus
Self-understanding	zelfbegrip	Thesaurus
Care	zorg	Van Dale
Self-care	zelfzorg	Dutch linguistic rules

Gentleness	zachtheid	Van Dale
Self-gentleness	zelfzachtheid	Dutch linguistic rules
Forgiveness	vergeving	Van Dale
Self-forgiveness	zelfvergeving	Dutch linguistic rules
Support	ondersteuning	Thesaurus
Self-support	zelfondersteuning	Dutch linguistic rules
Patience	geduld	Van Dale
Self-patience	zelfgeduld	Dutch linguistic rules
Warmth	warmte	Van Dale
Self-Warmth	zelfwarmte	Dutch linguistic rules
Acceptance	acceptatie	Thesaurus
Self-acceptance	zelfacceptatie	Dutch linguistic rules

Step Two: 'Judge Rating Phase'

The second step is the 'Judge Rating Phase' (Pennebaker et al., 2007). Here, the LIWC paper describes that three to four 'judges' or experts in the field check the list for the 'goodness of fit'. In the case of this thesis, the list was evaluated by the first and second supervisors and one other expert in the field of Positive Psychology to keep it feasible but still generate intersubjectivity in evaluating it.

For this step to be executed reliably, an instruction manual was created. This manual provided the definitions for SC and SCr, followed by written instructions on assessing each word and adding new ones. If one of the 'experts' indicated 'NO' for a particular synonym, this one was to be removed. The instructions for the Judge Rating Phase are described below. Also, Table 2 provides an example of part of the lexicon as it was shown to the experts.

"For this phase of the research, it is required that the lexicon created for self-compassion and self-criticism is checked by 'experts' to gain intersubjectivity and reliability.

The language central to this research is Dutch, meaning that the lexicon should be judged in Dutch (the English translation is added for the purpose of later publication). Below, there are four tables presented. Tables 1-3 each present synonyms for one separate facet of Self-compassion (Mindfulness, Self-kindness, and common humanity). The fourth table represents synonyms for self-criticism. Please decide per synonym whether you believe it belongs to the lexicon. Only check the box when you consider the synonym NOT to belong. Also, it is possible to write a remark. However, this is not mandatory. Finally, if you feel synonyms or concepts are still missing, please add them to the (fitting) table."

Table 2.

Synonyms for the facet of self-compassion: self-kindness (or encompassing the whole concept)

Synonyms	**Expert remarks**	**Check box if <u>NO</u>**
Mededogen		☐
Vriendelijkheid		☐
Zelfvriendelijkheid		☐
Begrip		☐
Zelfbegrip		☐
Zorg		☐
Zelfzorg		☐
Zachtheid		☐
Zelfzachtheid		☐

Following the experts' feedback, the lexicon was refined and made ready for testing using the developed corpus. See Appendix B for the complete instruction manual and the initial lexicon pre-judgement.

Step Five: 'Psychometric Evaluation'

The fifth step is the 'Psychometric Evaluation' (Pennebaker et al., 2007). This step is about validating the lexicon created to see if we are accurately measuring the construct of SC

and SCr. Our primary research question guides this evaluation: “In what way do expert perspectives and current literature contribute to the formulation of a valid and reliable lexicon intended to assess Sc and SCr in participants?”. Thus, our objective is to ascertain the validity and reliability of the measurements obtained from the lexicon.

The LIWC paper focuses more on the consistency and coherence between words, thus focusing on reliability analysis (Boyd et al., 2022). This thesis employs convergent validity, a subtype of construct validity (Abma et al., 2007), as part of its psychometric evaluation. To elaborate, the previous was done by initially testing the first two hypotheses. Moreover, two separate scores were created, one for the percentage of words that refer to SC and one for SCr. Next, test-retest reliability assessments were conducted on the lexical data to determine the consistency of the measures, which is essential for validating psychological instruments (Streiner & Kottner, 2014). The prior analysis involved assessing Pearson correlation coefficients between the pre-and post-measures of SC and SCr within the lexical data to answer hypotheses five and six.

Furthermore, to evaluate convergent validity, the correlations between the scores derived from the lexicon and those from established measures of SC, such as Neff’s SCS (2011), were examined to answer hypotheses three and four. In particular, the pre-measures between the lexicon and the SCS data were observed closely and reported in the results and discussion. The reason for this was that the interview data was presumed to be ‘naif’ in the sense that it was more naturally spontaneous and unaffected by certain words, sentences or nuances learned in the MBCL intervention.

Moreover, a strong connection between these measures would suggest that the lexicon accurately reflects the constructs of SC and SCr as initially intended. Neff (2011) asserts that self-compassion (SC) and self-criticism (SCr) represent opposing levels on a continuum, emphasising the necessity for accurate measurement of both constructs. Moreover, Reas et al.

(2010) emphasise the significance of convergent validity in validating that psychological assessments adequately reflect their intended constructs. Consequently, strong correlations with established indicators such as the SCS would substantiate the validity of our lexicon.

Additionally, paired t-tests were executed to compare the two means of the paired samples and test the hypotheses. This procedure allows for assessing ‘sensitivity to Change’, which was explained earlier and can detect meaningful changes over time (Fok & Henry, 2015). Despite other acceptable psychometric qualities, measurements that are insensitive to change may fail to detect the actual effects of an intervention, such as the MBCL, because sensitivity to change mediates detection (Fok & Henry, 2015). Additionally, focusing on this sensitivity is especially helpful in small sample studies, which tend to have low statistical power due to the sample size, to detect actual effects (Fok & Henry, 2015). In conclusion, these evaluations indirectly assess the validity of the lexicon by examining changes in SC and SCr.

Qualitative Analyses of a Single Interview

A qualitative single case analysis (N = 1) was performed to obtain a more profound understanding of the constructs of SC and SCr. This was implemented due to a prevalent issue in text mining, which involves the 'overlooking' of contexts in which particular words are utilised, impacting the results' accuracy and reliability (Iliev et al., 2014b). Furthermore, the prior analysis involved selecting a single participant and examining their pre- and post-interview responses by assessing the frequency of lexicon words mentioned, the total word count of the interview, and, most importantly, the contexts in which the words were utilized, establishing a thorough understanding of the participants' narrative. Subsequently, a reflective rapport was analysed using the interview findings, meticulously considering the subject's narrative and the employed lexicon, which was then analysed in conjunction with academic literature.

Additionally, the participants' individual self-report scores of the SCS were integrated into the qualitative analyses. This was to see whether our analyses aligned with the self-report measures on the established SCS. To this aim, the means and standard deviations of the Dutch version of the SCS, which include a 7-point likert scale, are derived from the paper by Raes et al. (2010). These findings were integrated into the qualitative analyses to compare the self-report measures of Participant 108 to the norm scores.

Lastly, demographic data was not disclosed to protect the participant's privacy. The rationale for selecting this specific subject is that the lexicon scores, measured within the interview, provided the opposite of what we had expected. Consequently, offering valuable material for qualitative analysis. In the result section, the findings of the analysis were revealed.

Statistical Analyses

Orange

To evaluate the developed lexicon utilising text-mining techniques, a program named 'Orange' was employed. This is a graphical visualisation tool capable of doing various text-mining analyses (Thakur & Kumar, 2020). The steps taken within Orange to incorporate the lexicon and analyse the interviews are described below.

First, the corpus was created in Orange. Meaning that the interview data was imported into the program. Second, the options for pre-processing the text were carefully installed. The following options for pre-processing were selected: ‘Lowercase’, ‘Remove accents’, Word punctuation’, Wordnet Lemmatizer’, Stopwords’, Lexicon’, Numbers’ and ‘Regexp’. Here, the lexicon was added so that the program was able to search the interviews for the words included on the list. After this, the option ‘word cloud’ was used, which presented words of the lexicon within the interviews and how often each was presented in total. After this, the ‘word cloud’ was used for each interview separately and also displayed how many and which

words were counted based on the lexicon terms and for each interview. This information enabled the calculation of scores (percentages) for SC and SCr, facilitating further analysis.

RStudio

Further analysis was done using the program ‘Rstudio’ (RStudio Team, 2022), which is a statistical software program that can perform various statistical analyses (Kronthaler & Zöllner, 2020). the following R packages were used: “haven” (Wickham et al., 2023), dplyr (Wickham et al., 2023), tidyr (Wickham et al., 2023), and ggplot2 (Wickham, 2016). Further steps of the statistical analyses that were performed are described below.

The descriptive statistics of the lexicon data were compiled. Indicating the mean and standard deviation of the SC data and SCr data for the pre-and post-measures (Table 4). Second, the first two hypotheses were tested using the previously described paired t-tests. Third, to test hypotheses five and six, correlations were computed using the Pearson correlation coefficient to evaluate the relationships between SC and SCr and their pre- and post-measures. Fourth, to test hypotheses three and four, Pearson correlation was employed to assess the relationships between the lexicon indicators and the participants' pre- and post-scores on the SCS created by Neff (2011) to evaluate the lexicon's validity. Lastly, Rstudio was also used to obtain graphs and figures of the results of the analyses.

Results

This results section will initially give the findings from step 2: 'The Judge Rating Phase', encompassing the complete lexicon, with a detailed discussion on the modifications made to the lexicon. The outcomes of the statistical analysis are thereafter presented, delineated, and exemplified by figures and tables. Finally, the results of the qualitative analysis were presented.

Post-Judge Rating Phase

Following the experts' completion of their examination of the preliminary list of synonyms for Self-compassion (SC) and Self-criticism (SCr) (refer to Appendix B), various adjustments were made to refine the lexicon. Initially, it was determined to utilise only single words rather than employing sentences composed of many words. This decision was made because data analysis would be less complicated and more practical.

Secondly, since a section of the lexicon was translated from English—given that numerous significant SC and SCr terms originated from English literature—in certain instances, the translation was exceedingly literal; however, the resultant terms, while grammatically correct, were not commonly employed by people in general to articulate SC or SCr. In English literature, a prevalent synonym for SC is 'self-kindness' (Neff, 2011). Nevertheless, the Dutch translation is 'zelf-liefde'; yet, according to the Dutch experts, the term "zelf" (“self” English translation) is infrequently employed, as they prefer alternative terms or expressions that align with this concept. For this reason, most words containing “Zelf” were omitted or replaced with another synonym related to the initial word, such as “self-kindness”.

Thirdly, experts observed that word lemmatisation is necessary. This involves offering various variations of the words in the list, as individuals may utilise particular words in diverse ways. Consequently, numerous words were incorporated into the lexicon.

Ultimately, the 'experts' identified synonyms to be removed from the lexicon and were authorised to propose new terms, resulting in the final composition of the lexicon provided in Table 3. Additionally, all newly added or revised synonyms are indicated in the table with an asterisk (*). There are a total of 191 terms for the construct SC and 144 terms for SCr. Furthermore, the lexicon includes additional linguistic categories such as verbs, nouns, stems,

and others. This was intended to integrate lemmatisation. This indicates that multiple variants of the synonyms were generated from the data.

Table 3.

Lexicon post judgement phase: SC (all facets) & SCr

Synonyms SC	**Synonyms SCr**
Mededogen	Oordeel
Vriendelijkheid	Oordelend*
Vriendelijk*	Kritisch*
Begrip*	Zelfkritisch*
Zelfbegrip*	Kritiek
Begripvol*	Zelfkritiek
Zorg	Lastigheden*
Zelfzorg	Lastigheid*
Zachtheid	Lastig*
Zacht*	Hardheid
Vergeving	Zelfhardheid
Zelfvergeving	Verharding*
Geduld	Verhard*
Warmte	Keihard*
Warm*	Hard*
Troost	Schuld
zelftroost	Zelfschuld
Aanmoediging	Schuldig
Veerkracht	Verwaarlozing
Kracht	Zelfverwaarlozing

Sterk*	Verwaarlozen*
Sterkte*	Verwaarloos*
Empathie	Schaamte
Empatisch*	Schaam*
Tolerantie	Veroordeling
Zelftolerantie	Vooroordeel*
Geduldig	Veroordeel*
Goedkeuring*	Straf
Zelfliefde	Straffend*
Eigenliefde*	Zelfstraf
Liefde*	Gestraft*
Aardig*	Twijfel
Eigenwaarde*	Twijfelt*
Waarde*	Zelftwijfel
Waardes*	Walging
Waarden*	Zelfwalging
Waarderen	Walg*
Zelfwaarde	Walgen*
Zelfcompassie*	Walgt*
Compassie*	Verwijt
Compassievol*	Verweet*
Mildheid*	Zelfverwijt
Mild*	Onvergeeflijk
Zelfzachtheid*	Negativiteit
Vergeven	Negatief*

Vergeef*	Zelfnegativiteit
Vergeeft*	Foutzoeken
Ondersteunen*	Fout*
Ondersteun*	Ontoereikendheid*
Ondersteuning	Ontoereikend*
Zelfondersteuning	Intolerant
Steun*	Ongeduldig
Omarmen	Ongeduld*
Omarm*	Falen*
Begrijpen	Faal*
Begrijp*	Faalt*
Tolereren	Faalde*
Tolereer*	Gefaald*
Tolerant	Beneden*
Genade*	Streng*
Mindfulness	Zelfverlagend
Mindful*	Verlagen*
Niet-oordelend	Verlaagd*
Acceptatie	Verlaag*
Zelfacceptatie	Zelfonderzoek
Accepteren*	Wantrouwen
Accepteer*	Afkeuring
Accepteert*	Afkeuren*
Geaccepteerd*	Keur*
Niet-kritisch	Zelfevaluatie

Kalmeren	Zelfcommentaar
Kalm*	Commentaar*
Kalmte*	Bezwaar
Lief*	Zelfaanval
Zorgzaam	Aanval*
Menselijkheid	Zelfbeschuldiging
Menselijk	Beschuldiging
Gedeeld	Beschuldig*
Delen	Perfect*
Barmhartigheid	Bekritiseren*
Barmhartig*	Bekritiseer*
Aanvaard	Bekritiserend*
Aanvaarden*	Bekritiseerd*
Alledaags	Zelfprojectie*
Dagelijks	Projecteren
Vertrouwd	Dom*
Natuurlijk	Onhandig*
Heersend	Stom*
Wijdverspreid	Idioot*
Typisch	Moeilijk*
Universeel	Moeilijkheid*
Verbondenheid	Moeilijkheden*
Verbonden*	Moeizaam*
Verbinding	Piekeren*
Kwetsbaarheid*	Pieker*

Kwetsbaar*	Piekeraar*
Solidariteit	Perfectionistisch*
Best*	Verwachting*
Goed*	Verwacht*
Houden*	Wegcijferen*
Doormaken*	Cijfer*
Meemaken*	Onzeker*
Ervaren*	Minderwaardig*
Gemeenschap*	Projecteer*
Gemeenschappelijkheid	Afkeur*
Gemeenschappelijk	Afgekeurd*
Iedereen	Aarzel*
Vertrouwen*	Aarzelend
Vertrouw*	Aarzelen*
Ademruimte*	Geaarzeld*
Ademen*	Verwijten*
Break*	Verwijt
Pauze*	Verweten*
Herstel*	Keuren*
Herstellen*	Sloom*
Terugkomen*	Slomer*
Mogen*	Traag*
Mag	Trager*
Open*	Lui*
Vooruit*	Boos*

Zingeving*	Gesloten*
Strijd*	Geslotenheid*
Gestreden*	Beoordelen*
Strijden*	Beoordeling*
Positiviteit*	Beoordeel*
Relativeren*	Koud*
Relativeer*	Kil*
Relativeert*	Strikt*
Gerelativeerd*	Precies*
Omgaan*	Nauwkeurig*
Harmonie	Nauwgezet*
Bewsutwording*	Consequent*
Bewustworden*	Scherp*
Bewustheid*	Onbuigzaam*
Aanmoedigen*	Veeleisend*
Liefdevol*	Moeizaam*
Verdragen*	Kieskeurig*
Verdraag*	Onvrede*
Verdraagt*	Ontevreden*
Vrede*	Afwijzing*
Toelaten*	Afwijzen*
Toegelaten*	Afkraken*
Emoties*	Kraak*
Emotie*	Kraken*
Voelen*	

Laten*
Toestaan*
Toegestaan*
Goedkeuren*
Gedogen*
Dogen*
Ondervinden*
Ondervindt*
Ondervind*
Aanvoelen*
Beseffen*
Besef*
Gewaarworden*
Gewaarwording*
Trotseren*
Trots*
Verduren*
Overgave*
Overgeven*
Verzorging*
Verzorg*
Verzorgen*
Aandacht*
Inzicht*
Inzichtelijk*

Rust*
Rustpunt*
Rusten*
Gerust*
Geruststellen*
Geruststelling*
Helpen*
Hulp*
Genoeg*
Reflecteren*
Gereflecteerd*
Reflecteer*
Reflecteert*
Reflectie*
Loslaten*
Los*
Pijn*
Geleden*
Lijden*
Leed*
Verzachten*

Test-Retest Reliability

After completing step two, the lexicon was used for further research. Initially, scores (percentages) were computed by counting the number of lexicon words present in the

interview, dividing by the total number of words in the transcribed interview, and multiplying by 100. The results are displayed in Appendix C, Table C1.

Furthermore, the test re-test reliability measures were executed on the lexicon data to evaluate the consistency of the measures over time, which is essential for establishing the validity of psychological instruments (Streiner & Kottner, 2014). The previous was done by testing Pearson correlation coefficients between the pre and post-measures of SC and SCr of the lexicon data. Table 4 presents the results of these analyses.

The results to hypotheses five and six, revealed a significant correlation between SC prior to the intervention and subsequent measurements, indicating moderate consistency over time. In contrast, the association between pre-intervention measures of SCr and post-intervention indicated minimal stability. Moreover, there were no significant associations detected between pre-intervention measures of SC and pre-intervention measures of SCr, nor between pre-intervention measures of SC and post-intervention measures of SCr, suggesting that the measure of SC and SCr do not consistently correlate across evaluations nor within. Additionally, the association between post-intervention measures of SC and pre-intervention measures and between post-intervention measures of SC and post-measures of SCr all indicate no significant correlations.

Table 4

A Correlation Matrix Exhibiting the Associations Between the Lexicon Variables.

		Pre_SC	Post_SC	Pre_SCr	Post_SCr
Pre_SC	*r*	1	.573	-.016	-.280
	p		.003**	.942	.185

Post_SC	*r*	1	.066	-.208
	p		.759	.330
Pre_SCr	*r*		1	-.090
	p			.677
Post_SCr	*r*			1
	p			

Note. * = *significance,* *p < .05, **, p < .01, *** p < .001

Sensitivity to Change

Descriptive Statistics

In the following section, descriptive statistics of the lexicon data were presented in Table 5 and illustrated in Figure 1. Following this, paired t-tests have been conducted to compare the paired samples' means to evaluate the first two hypotheses and are also presented in Table 5. As previously explained, this approach facilitates the evaluation of 'sensitivity to change,' and can identify meaningful changes over time (Fok & Henry, 2015). As presented in Table 5, it can be noted that the mean scores are low, ranging from .005 to .034. This may suggest that the baseline levels of SC and SCr among the participants were already low. Furthermore, the change in mean scores for SC from 0.034 prior to the intervention to 0.032 subsequent to the intervention contradicts the anticipated result of hypothesis one, which predicted an increase in SC following the intervention. Finally, the mean scores for SCr were consistently low, recording 0.009 pre-intervention and 0.005 post-intervention, demonstrating negligible variation.

These results are further exhibited in Figure 1. A box plot illustrates the comparison between SC and SCr pre- and post-measures. Moreover, it illustrates the distributions of

scores, emphasising outliers, the median, and the interquartile range. The boxplot indicates that the pre-intervention data exhibit low levels of SC and SCr, with subjects' SC scores concentrated at the lower end of the scale. Although minor changes in post-intervention scores are evident, the overall change in SC did not demonstrate a significant trend upwards. The median following the intervention SC scores is, in fact, lower than those recorded prior to the intervention, consistent with the previously indicated results. However, the SCr scores exhibit a slight decline post-intervention, indicating reduced variability and further substantiating the concept of negligible change throughout the intervention.

Table 5.

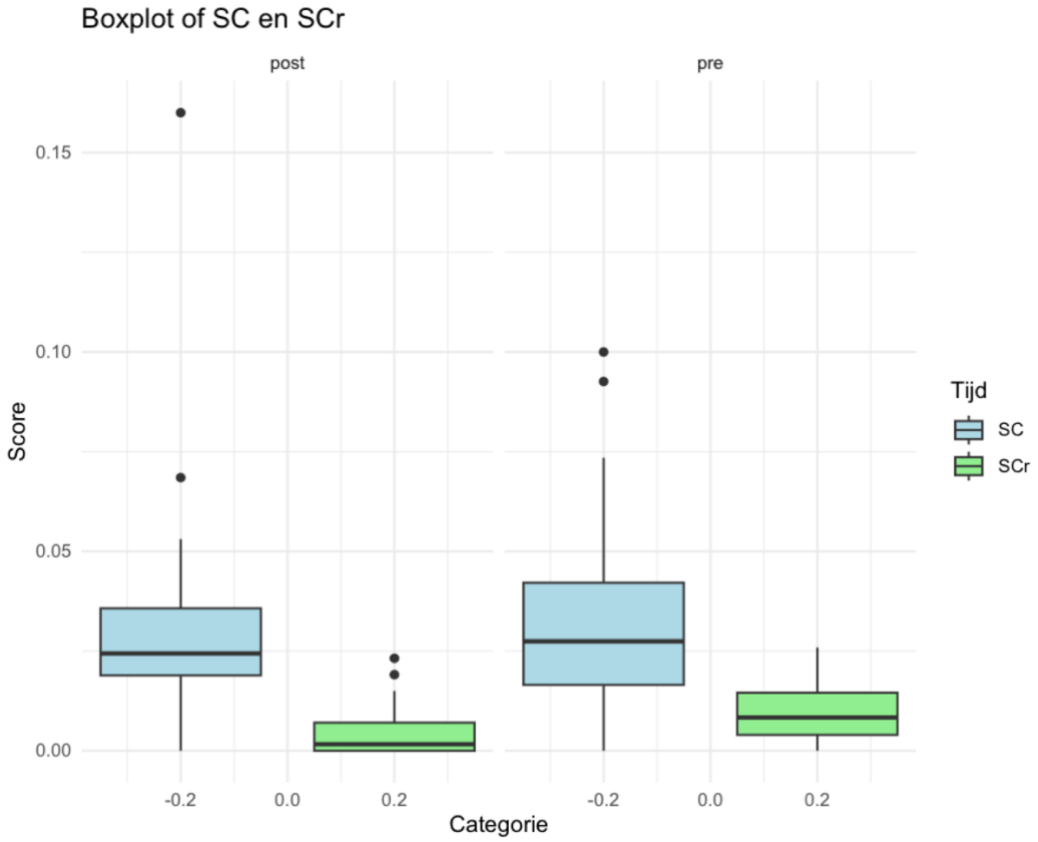
Descriptive statistics of lexicon data and results of the paired t-tests of the lexicon data.

	Pre_SC	**Post_SC**	**Pre_SCr**	**Post_SCr**
Mean	.034	.032	.009	.005
SD	.026	.032	.007	.007

Hypothesis	**Score Type**	**t-value**	**df**	**p-value**	**Mean Difference**	**95% Confidence Interval**
H1	SCr	2.037	23	.053	.0042	[-0.0000663, 0.0085]
H2	SC	.265	23	0.793	.001	[-0.00996, 0.01289]

Figure 1.

Presents a boxplot of the pre- and post-measures of SC and SCr.



Paired t-tests

First, hypothesis one (H1), stating that SCr would decrease from baseline to post (.004), was tested in a paired t-test. This showed a marginally significant decrease from baseline (m=.009) to post (m=.004) of 0.0042, as seen in Table 5.

Second, hypothesis two (H2) posited that SC would increase post-intervention compared to pre-intervention levels. The results indicate a lack of support for this hypothesis, evidenced by a t-value of .26549, (p = 0.793). Consequently, this indicates no substantial change in SC. The mean difference for SC was 0.001, indicating an absence of effect.

Construct Validity

To assess convergent validity, a subtype of construct validity (Abma et al., 2016), the SC and SCr scores of the lexicon were analysed using Pearson correlation coefficients alongside existing measures of SC, such as Neff's SCS (2011), to test hypotheses three and

four. Moreover, the pre-measures have been selected for reporting in this thesis. The post-intervention measures, including tables, figures and descriptions can be found in Appendix D.

Table 6 delineates the associations between the pre-intervention SC scores derived from the lexical data and the pre-intervention scores of the participants on the SCS scale, along with its distinct components. Furthermore, the correlations with self-kindness ($r = 0.086$, $p = 0.691$), common humanity ($r = 0.097$, $p = 0.652$), and mindfulness ($r = 0.017$, $p = 0.939$) indicate negligible and weak relationships. The correlation with the overall SCS score is poor ($r = 0.061$, $p = 0.779$). The association was visually depicted in a scatterplot in Figure 2.

Table 6.

Correlations with pre-SC scores and pre-SCS scores.

Variable pair	**Correlation**	**P-value**
SC - Self-kindness	*.086*	*.691*
SC - Common Humanity	*.097*	*.652*
SC - Mindfulness	*.017*	*.939*
SC - Self-judgement	*-.001*	*.998*
SC - Isolation	*-.037*	*.864*
SC - Over-identification	*-.079*	*.713*
SC - SCS_total	*.061*	*.779*

Figure 2.

This exhibits the correlation analyses between the pre- SC scores of the lexicon data and the total SCS data (T0)

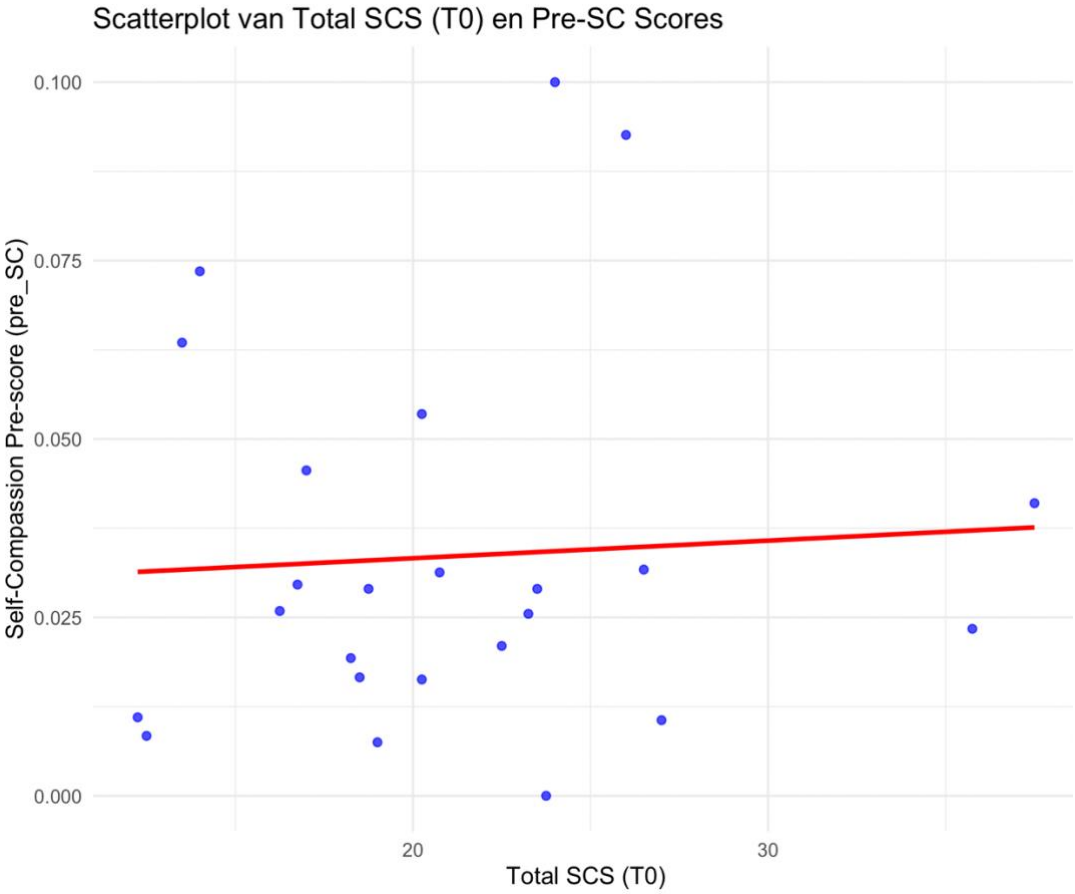


Table 7 illustrates the correlations between the pre-intervention SCr values and the pre-intervention SCS ratings. The connections with self-kindness (r = -0.074, p = 0.731), common humanity (r = -0.235, p = 0.269), and mindfulness (r = -0.221, p = 0.299) indicate weak and inconsequential interactions. The connection with the total SCS is modest (r = -0.112, p = 0.603). Figure 3 further elucidates the findings with a scatterplot depicting the correlation between pre-SCr scores and overall SCS scores prior to the intervention.

Table 7.

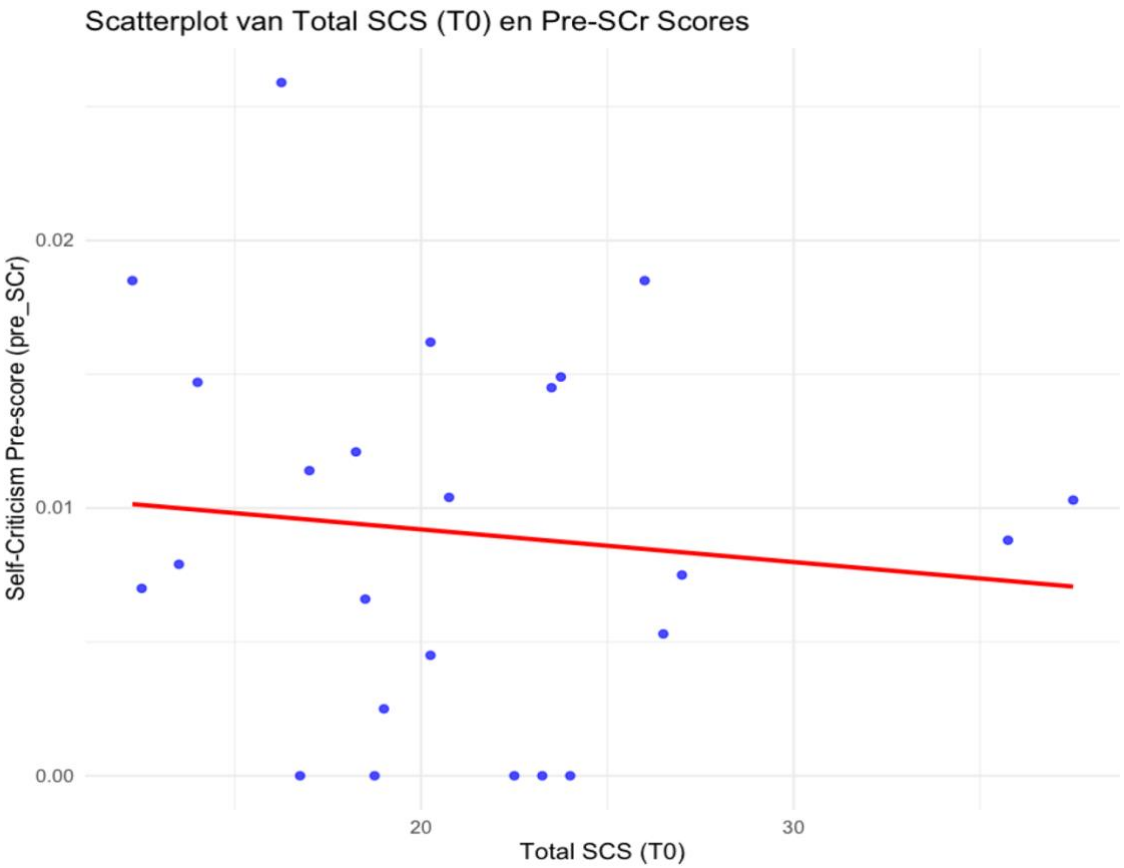
Correlations pre SCr scores and pre-SCS scores

Variable pair	**Correlation**	**P-value**
SCr - Self-kindness	*-.074*	*.731*
SCr -_Common Humanity	*-.235*	*.269*

SCr - Mindfulness	*-.221*	*.299*
SCr - Self-judgement	*.025*	*.908*
SCr - Isolation	*.117*	*.586*
SCr - Over-identification	*-.107*	*.618*
SCr - SCS_total	*-.112*	*.603*

Figure 3.

This exhibits the correlation analyses between the pre- SCr scores of the lexicon data and the total SCS data (T0)



Qualitative Analysis of Single Subject

The selected subject for the qualitative analyses was participant number 108 (Table 4). This individual was selected due to the observation that the SC scores decreased following the intervention, whereas the SCr scores increased thereafter. This unexpected outcome suggests that the participant interview data may yield valuable insights. Moreover, the pre- and post-interview was qualitatively analysed and discussed.

Pre-intervention Interview

The first interview consisted of 270 words. 8 words from the lexicon were found in the interview and all belonged to the SC category. The following words were: “sterk” (three times), “goed” (two times), “zingeving” (two times), and “Strijd” (one time). Thus, the calculated percentage for participant 108 for pre-intervention SC was 2.96% while the score for SCr was zero. Additionally, the participant’s scores of the SCS were presented in Table 8.

While assessing the transcribed interview, it was noticeable that the participant frequently employed the term “Sterk” (Strong) during the interview to express how strongly she felt about her personal needs. The recurrent employment of this term in this context may signify an acknowledgement of one's capacity to navigate the complexities of life, which can be an essential component of self-acceptance. This is due to its emphasis on recognising one's ability and personal needs, hence, one's worth (J. W. Zhang et al., 2019). Moreover, J. W. Zhang et al. (2019) emphasise the important relationship between self-acceptance and SC as they stated that SC predicts self-acceptance as well as acceptance of others, highlighting the importance of focusing on this aspect.

Next to this, the participant also mentioned “zingeving” (meaning) multiple times, which indicates a possible desire for purpose and meaning in life. This became apparent by the quote: ““ik ben altijd heel erg aan het zoeken naar dingen van hoe kan ik dingen doen die waar ik zelf me goed bij voel en waar ik een stuk zingeving bij voel.”, which translates to: "I

am constantly searching for activities that will allow me to find meaning in my life.". Additionally, the participant denotes that she has not yet integrated this in her life. Moreover, finding purpose or experiencing this is a fundamental aspect of SC as indicated by Neff (2011). Other research found that SC facilitates experiencing purpose in life (Suh & Chong, 2021). Knowing this, it can be carefully implied that the participant's lack of current experiencing meaning could reflect a lower sense of SC.

Interestingly, measuring this word with the lexicon implies a higher sense of SC, while it became apparent that, in this context, it reflects lower SC. This is because the participants emphasise that there is a desire or search for meaning ("zingeving"), but the subject has not managed to integrate this into her life yet. In this case, it is seemingly important to understand the context as well as the narrative of the story to see whether the word reflects SC or SCr.

Another lexicon word that was mentioned was the term "Strijd" ("conflict"). While observing the context, it became apparent that the participant indicated an internal conflict, which led to experiencing difficulties in finding her place in society. This became apparent by the quote: "ik ben altijd heel erg aan het zoeken naar dingen van hoe kan ik dingen doen die waar ik zelf me goed bij voel en waar ik een stuk zingeving bij voel en hoe doe je doe ik dat in een maatschappelijke context", which translates to the participant trying to find meaning in the things the subject does and how he/she can integrate this in societal context. The participant elaborates that she seeks ways to "integrate both worlds." This underscores a disparity between self-assessed strengths and the participant's actual experiences, potentially indicating an identity conflict.

In line with this, the participant struggled to provide a concrete answer to how she sees herself, which could indicate a struggle with defining personal identity. She mentioned: " ik zie mezelf denk ik als iemand die toch zijn leven als een ingewikkelde set van keuzes ziet", which translates to: I see myself as someone who views life as a set of complicated choices".

Neff (2011) asserts that internal conflict corresponds with the notion that self-compassion involves acknowledging one's difficulties or 'suffering' within the larger context of human existence. Eventhough the participant acknowledges both inner strengths and suffering, which is in line with the argument by Neff (2011), the subject encounters difficulties in incorporating these strengths into daily life, which possibly reflects a lower level of SC. This indicates that the word “Strijd” should also be viewed in accordance with the context.In conclusion, the participant’s narrative reveals a possible internal conflict regarding their identity, meaning and emotional state. The words "sterk" and "goed" may indicate that although they are aware of their strengths, they also have trouble combining them with their search for purpose in life, as became apparent by the lexicon words: “Zingeving” and “Strijd”. A moderate SC score may be reflected in this duality. However, the complex nature of SC is demonstrated by their attempts to match their personal beliefs with those of society, since people may be both capable of overcoming life's challenges and facing uncertainty or self-doubt.

Post-intervention Interview

The post-intervention interview of participant 108 contained 206 words, of which three words of the lexicon were found. These words were: “precies” (precise), “kritisch” (critical), and “fout” (wrong/mistake). All of these words pertain to the SCr category of the lexicon. Thus, reflecting her post-scores of 0 per cent for SC and consequently a decrease from pre-intervention to post-intervention, and an increase in SCr from 0 percent pre-intervention to 1.46% thereafter.

Interestingly, the participant uses these words in an opposite context of what the current lexicon measures implied. To elaborate, the subject mentioned: “… maar dat ik daarin een wat rustigere houding naar mezelf toe heb ontwikkeld. In ieder geval minder, minder

kritisch.”. This translates to the subject explaining to have a less critical view of the self, which highlights the context and the narrative of the subject. Also, it is noticeable that the participant uses the word “minder” (less) often as an indication of less issues of complaints. Moreover, words like “Less” or “more” create the context and the narrative of the story, and since the lexicon does not account for this, it could be that a high score of SC or SCr can be found, but that it actually measures the opposite.

This is the case for this participant, as a decrease in SC was found and an increase in SCr. The participant indicated to be less self-critical and less focused on ‘mistakes’, possibly reflecting a decrease in SCr rather than an increase. Additionally, the subject shows more signs of self-acceptance in the post-interview. This was derived from the quote: “dat ik daarin wel een mildere houding naar mezelf toe heb ontwikkeld.”, which translates to the subject having a milder view of oneself. It can be argued that having a milder view of oneself indicates higher levels of self-acceptance, which is a fundamental aspect of SC (Nef, 2011). Therefore, after analysing the interview, it can be expected the subject showed an increase in SC rather than a decrease.

In conclusion, it is seemingly important to assess the context in which particular words or expressions are used, as the context determines the story's narrative. After comparing the pre-intervention scores of the lexicon of participant 108 to a qualitative report of the pre-intervention interview, it can be argued that the SC scores of the lexicon could be in accordance with the interview despite the complex nature of SC. However, the lexicon measured zero per cent of SCr within the interview, while the participant showed uncertainty and self-doubt, which aligns with SCr (Potter et al., 2014). This thus reflects a potentially higher score in SCr than the lexicon measured.

Upon comparing the post scores of the lexicon with the post-intervention interview, it became evident that terms like “less” and “more” served as significant indicators of the

narrative context. The usage of “less” signified a reduction in complaints and SCr. In contrast, the lexicon recorded elevated levels of self-criticism, as it only assessed isolated terms such as “Kritisch” (critical) and “fout” (wrong/mistake). Therefore, an analysis of the interview suggests that the participant should exhibit an improvement in SC due to enhanced self-acceptance, alongside a decrease in SCr, as evidenced by the contextual observations.

Comparing Findings With Self-Report Measures SCS

In order to compare the qualitative findings to self-report measures of participant 108 on the SCS, the subject’s outcomes for each subscale have been presented in Table 8. Participant 108's self-report measures on the SCS indicate a rise in self-compassion, as the subscales of 'self-kindness,' 'common humanity,' and 'mindfulness' all demonstrate improvement from pre- to post-intervention. Furthermore, the subscales of ‘self-judgment’, ‘isolation’, and ‘over-identification’ show a decrease from pre to post-intervention. These results are contrary to the lexicon findings, which show a decrease in SC from pre to post-intervention and an increase in SCr. However, the SCS results align with the findings from the qualitative analysis, indicating the need for a closer examination of the lexicon.

Table 8.

This table presents the pre- and post SCS data of participant 108 as well as the normscores.

SCS Subscales	T0 Scores	T1 Scores	Mean (Dutch version)	SD
Self-kindness	3.00	4.75	4.73	1.05
Common humanity	3.25	4.75	4.71	1.15
Mindfulness	4.00	4.50	4.94	1.08
Self-judgement	5.75	3.50	3.30	1.16

Isolation	5.50	3.75	3.24	1.22
Over-identification	6.25	3.50	2.78	1.17

Note. *The Dutch version of the SCS was used and includes a 7-point likert scale (Raes et al., 2010). T0 = pre-intervention scores, T1 = post intervention scores. The mean and SD are derived from literature (Raes et al., 2010).*

Discussion

This thesis attempted to address a gap in the LIWC text mining literature about the notions of SC and SCr by creating and validating a comprehensive lexicon or dictionary conceptually based on part of the steps of the LIWC (Pennebaker et al., 2007) and beyond. This lexicon was intended to be a novel assessment tool for measuring SC and SCr in a text-based context. Further, this study aimed to answer the research questions: In what way do expert perspectives and current literature contribute to the formulation of a valid and reliable lexicon intended to assess SC and SCr in participants?, Are SC and SCr best represented on a unidimensional or bidimensional scale? And lastly: In what way do participants' narratives in qualitative interviews illustrate their experiences of SC and SCr prior to and after the MBCL intervention?

Summary of Findings

For H1, the mean frequency of lexicon words related to SCr was anticipated to be higher before the MBCL intervention than after. The results provided marginal support, with a lower post-intervention mean for SCr (mean = .005) than pre-intervention (mean = .009). However, H2, where the expectation was that the mean frequency of lexicon words associated with SC would be higher after the MBCL intervention, was not supported. For H3, it was posited that there would be a significant positive correlation between the SC score of the lexicon and the SCS scores; the results showed a weak and non-significant association (r =

.061, p =.779), and for H4, whereby it was expected there would be a significant negative correlation between the SCr scores of the lexicon and the SCS. Yet, a modest, non-significant relationship was observed (r = -.112, p = .603). Next, H5 and H6, aimed to assess the stability of the lexicon via test-retest reliability measures, showed a significant correlation for the SC lexicon (r = .573, p = .003), demonstrating consistent results over time, while SCr showed no significant correlation.

To answer the first RQ, it can be argued that the lexicon failed to accurately reflect SC and SCr and showed weak validity and reliability according to hypotheses testing, test-retest measures and the cross-correlations measures. Hence, it can be stated that the lexicon is not a valid and reliable measuring tool. Pertaining to the second RQ, it can be argued that SC and SCr function more effectively as independent dimensions rather than a unidimensional construct, as evidenced by the poor correlations between lexicon scores and established measures, coupled with a notable imbalance in the representation of the two constructs within the lexicon. Lastly, the key finding of the third RQ is that the lexicon failed to capture the nuances used by the participant to indicate SC and SCr as contextual understandings from participants' narratives illuminated the complex emotional landscapes surrounding SC and SCr.

Interpretation of the Findings

Sensitivity to Change

This thesis aimed to assess the validity of the lexicon by examining the first two hypotheses. The results did not fully support these assumptions, as the subjects demonstrated a marginally significant decrease in SCr scores and notable stability in SC levels following the intervention. Remarkably, the participant sample in the study by Schuling et al. (2016), from which we utilised a subsample, showed significant improvements in self-compassion as measured by the Self-Compassion Scale (Neff, 2011). Another remarkable finding is that the

nearly significant decrease in SCr aligns with the findings by Schuling et al. (2016); however, due to the results of the test-retest measures, which exhibit a lack of consistent correlation in SCr scores across different time points, it can be argued that the findings are not reliable.

A possible explanation for this discrepancy could be that the lexicon employed to evaluate SCr may be insufficiently sensitive to identify nuanced alterations in self-critical thoughts that may arise due to the intervention. Moreover, sensitivity to change is critical in capturing the nuances of emotional shifts, especially in small sample sizes (Fok & Henry, 2015). The weak correlation in the test-retest data ($r = -.09$, $p = .677$) suggests that the measure may inadequately capture transient fluctuations, highlighting a reliability issue that compromises the interpretation of the marginally significant decrease observed in the t-test ($p = .053$). These findings underscore the need to employ reliable and sensitive measurement instruments to evaluate psychological constructs effectively.

The Lexicon's inadequacy was possibly also demonstrated by the low mean scores (ranging from 0.05 to 0.34), suggesting that it inadequately captured the complexities of SC and SCr (Leary et al., 2007). Additionally, when assessing the mean scores in conjunction with comparable literature, a study by Ji and Raney (2020d) offers valuable insights. They created a self-transcendent emotion dictionary comprising 351 words categorised into six groups: awe-oriented, gratitude-oriented, elevation-oriented, admiration-oriented, hope-oriented, and general inspiration-oriented words (Ji & Raney, 2020d). Furthermore, in creating the dictionary, they followed all the previously outlined processes of the LIWC (Ji & Raney, 2020d). Their corpus was substantially larger ($n = 3,024$) than the corpus utilised for this thesis ($n = 24$). Their mean scores across the six categories range from 0.31 to 0.67. Given that mean scores for text-mining methodologies are generally low (Ji & Raney, 2020d; Boyd et al., 2022), the scores from the study by Ji and Raney (2020d) are significantly higher

than those obtained from the lexicon. This may suggest that the lexicon fails to appropriately assess the constructs of SC and SCr and its sensitivity to change.

Another argument for the findings regarding the mean scores is that the interview data was presumably not a 'spontaneous' text, which means that participants provided input for the data without prompts or suggestions. In this case, the interviews were led by questions like: "How do you experience yourself?" which can be considered prompting (Salecha et al., 2024), as it can trigger biases in participants, such as the social desirability bias, to provide socially fitting answers (Salecha et al., 2024). Consequently, it could be expected that the means of the lexicon can be considered higher than they would be if the data were to be 'spontaneous', further highlighting the lexicon's inadequacy.

Construct Validity

This thesis emphasised the importance of assessing construct validity by testing hypotheses three and four. However, the analyses revealed inadequate results, as the weak associations between the established SCS measures and the lexicon imply that the lexicon may not comprehensively encapsulate the intended constructs (Westen & Rosenthal, 2003).

A possible reason for this could be due to inadequate lexicon development, as the lexicon may have been constructed with a limited array of terms. This inadequately captures the emotional experiences associated with SC and SCr. According to Newton (2012), a limited focus may result in the oversimplification of complicated concepts, thus leading to incomplete assessments that fail to represent individuals' psychological states adequately. This limitation speaks to the issue of semantics in that the lexical approach may overlook the rich meanings and emotional contexts conveyed through language, ultimately affecting the validity of the constructs measured.

Building on the latter, as the lexicon concentrated exclusively on individual words rather than sentences or expressions, the context and nuances of the words' usage were thereby overlooked. Such semantic processing limitations can lead to emotional state misinterpretations, as meanings can shift dramatically based on the surrounding context (Holtgraves, 2015). Consequently, false positives and negatives were identified from the data, undermining construct validity (Stavrianou et al., 2007). This finding was also found in the qualitative analysis of the participant's interview subject, whereby the analysis revealed that the context in which specific lexicon terms were used significantly influenced their meanings. While the subject expressed increased SC and reduced SCr, the lexicon measured the opposite. Thus, it failed to capture this shift because it focused on isolated word usage rather than the narrative context. This finding aligns with assertions by Jusoh (2018), who emphasised that ignoring context in text analysis contributes to inaccuracies, thus advocating for methodologies that consider contextual nuances to increase construct validity.

Choice to Only Present Pre-Intervention Measures. For this thesis, the choice was made only to present the pre-intervention measures that were correlated alongside the pre-intervention measures of the SCS by Neff (2011). The post measures are to be found in Appendix D. To elaborate; this choice was based on the argument that the MBCL intervention could have affected the language of its participants in the sense that the subjects 'learn' to use particular words, often used in therapy, to describe or express their emotional or psychological states. Furthermore, a study by Grazzani and Ornaghi (2011) examined whether active engagement in discussing emotional states considerably enhances the comprehension of emotional terms and improves emotional understanding. Their results indicated that the experimental group surpassed the control group in grasping inner-state language and comprehending emotions (Grazzani & Ornaghi, 2011). Thus, as the MBCL intervention could

have a similar effect on the results of the post-intervention measures, the choice was made only to present the pre-intervention measures as these are presumed to be ‘naif’ and unaffected by the intervention.

Test-Retest Reliability

For hypotheses five and six, test-retest measures were executed, whereby the variables of the lexicon were correlated. The results showed that the pre-measures of SC presented a significant correlation to the post-measures of SC, while the pre-SCr scores with the post-SCr scores did not. This may reflect a lack of attention towards the terms of SCr, resulting in an incomplete set of terms that do not fully measure the whole construct. Similarly, in light of the ‘Judgement Rating Phase’, it became apparent that there was a substantial difference in the length of the SC and SCr lists. While the list for SCr comprised 144 terms, The list for SC included 191 terms, possibly indicating a greater emphasis on SC than SCr.

Building on the observation that the test-retest measures for SC remained constant over time, whereas SCr did not, this may be attributed to the Dutch language vocabulary. Moreover, Moors et al. (2012) examined affective research in the Dutch language, analysing 4,300 Dutch terms. The study indicated that positive affect words were more prevalent and shorter than negative affect words and that positive affect words were acquired earlier in childhood (Moors et al., 2012). Consequently, it can be posited that the terminology for SC is utilised more frequently in general than the terminology for SCr, which may elucidate the results of the test-retest measurements.

Limitations & Future Research

This thesis offers insights into measuring SC and SCr via its proposed lexicon. Nonetheless, particular limitations necessitate examination, as they may influence the

findings' validity, reliability, and generalizability. The limitations are categorised, emphasising critical areas for enhancement and establishing a foundation for future research. The first limitation pertains to adherence to the Language Inquiry and Word Count (LIWC) procedures. Although LIWC is a well-regarded instrument for linguistic analysis, rigorous compliance with its protocols is essential for accurately evaluating emotional content (Boyd et al., 2022). This thesis created the lexicon without complete adherence to all LIWC criteria, potentially leading to a less nuanced comprehension of language use related to SC and SCr. Consequently, future research may gain from utilising advanced machine-learning language models that more adeptly identify nuanced grammatical patterns, contextual connections, and emotional subtleties than standard lexicons (Devlin et al., 2019). Models like BERT and GPT-3 offer more nuanced assessments, reflecting contextual implications that LIWC may overlook.

The lack of an internal consistency analysis for the newly constructed lexicon further constrains the conclusions of this thesis. The dependability of the lexicon cannot be sufficiently evaluated without metrics of internal consistency, such as Cronbach's alpha (Agbo, 2010). Internal consistency analyses are crucial for assessing the coherence of items inside a scale, confirming that they measure the same underlying notion. Subsequent research should integrate reliability evaluations to ensure that the lexicon items consistently represent the constructs of interest. This would augment the credibility and the scientific validity of the findings presented in this research.

Another critical limitation is the sample size (N = 24), which may not provide adequate statistical power to identify subtle variations in SC and SCr. A limited sample size compromises the generalizability of findings to the broader population, as smaller samples often exhibit more significant variability in results and reduce the precision of effect size estimates (Cohen, 1992). Insufficient statistical power can result in Type II errors, where

actual effects go undetected. Future research should prioritise larger sample sizes to enhance the reliability of findings and improve the capacity to identify meaningful changes in emotional constructs. The choice to create two distinct lexicons, one for SC and another for SCr, imposes constraints on the overall coherence of the results. The thesis introduces difficulties that may obstruct the interpretation of outcomes by maintaining two distinct scales. The dependence on a two-dimensional paradigm necessitates thorough consideration, as it presupposes that SC and SCr operate on separate continua, perhaps failing to accurately represent the interaction between these constructs (Neff, 2011). Subsequent research would be enhanced by integrating these lexicons into a cohesive model. A unified lexicon could enhance the comprehension of self-compassion and self-criticism, enabling a more precise evaluation of their interplay and impact on emotional well-being.

Practical Implications

This study's findings have multiple practical implications for evaluating and improving SC and SCr in therapeutic and research contexts. Firstly, the test-retest measures demonstrated significant findings for SC, indicating stable measurements over time. Despite the limited sample size, it can be cautiously inferred that the lexicon may be a dependable instrument for assessing SC. Furthermore, this stability enables practitioners to track alterations in SC over time, hence aiding in the assessment of intervention efficacy, especially in therapeutic settings like the MBCL intervention (Germer & Neff, 2013).

However, the discovery that the test-retest measures exhibited instability for SCr over time implies a necessity for further development of the terminology for SCr. In order to improve the measurement of SCr, researchers and practitioners could consider expanding the lexicon to include a broader range of terms that encapsulate the various aspects of SCr. In order to help ensure the lexicon accurately reflects the construct of SCr, qualitative insights

from participants' narratives could be integrated into its construction (Neff, 2011), providing a more nuanced understanding of clients' self-critical experiences.

Furthermore, the findings indicate that cultural and language elements must be considered when creating and implementing psychological instruments. As evidenced by the Dutch language context in this study, The familiarity and frequency of specific terminology significantly influence how individuals connect with and express their experiences of SC and SCr (Moors et al., 2012). Consequently, practitioners need to recognise these contextual aspects to enhance the usability and efficacy of evaluation techniques across varied groups.

Conclusion

This thesis created and evaluated an innovative dictionary for assessing SC and SCr, focusing on three principal research objectives. The results indicated that the lexicon exhibited neither validity nor reliability, evidenced by modest correlations with established measures such as the SCS. Moreover, SC and SCr ought to be regarded as distinct concepts due to their substantial lexical representation disparities. Qualitative narratives elucidated the intricacies of participants' experiences that the dictionary failed to encompass, highlighting the necessity for context-specific terminology to enhance treatment efficacy. Incorporating these insights may improve the comprehension of SC and SCr and facilitate future studies in this domain.

References

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). https://doi.org/10.1186/s13104-016-2034-2

Agbo, A. A. (2010). Cronbach's Alpha: Review of Limitations and associated Recommendations. *Journal of Psychology in Africa, 20*(2), 233–239. https://doi.org/10.1080/14330237.2010.10820371

Baer, R. A., Lykins, E. L., & Peters, J. R. (2012). Mindfulness and self-compassion as predictors of psychological wellbeing in long-term meditators and matched nonmeditators. *The Journal of Positive Psychology, 7*(3), 230–238. https://doi.org/10.1080/17439760.2012.674548

Barnard, L. K., & Curry, J. F. (2011). Self-Compassion: Conceptualizations, Correlates, & Interventions. *Review of General Psychology*, *15*(4), 289–303. https://doi.org/10.1037/a0025754

Bayir, A., & Lomas, T. (2016). Difficulties generating self-compassion: An interpretative phenomenological analysis. *The Journal of Happiness & Well-Being, 4*(1), 15–33. http://www.journalofhappiness.net/3.pdf

Bittermann, A., & Fischer, A. (2023). "Text Mining in Psychology." *Zeitschrift Für Psychologie, 231*(1), 79–80. https://doi.org/10.1027/2151-2604/a000514

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best Practices for developing and Validating scales for health, Social, and Behavioral Research: A primer. *Frontiers in Public Health, 6*. https://doi.org/10.3389/fpubh.2018.00149

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *University of Texas at Austin, TX*. https://www.liwc.app

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Cox, B. J., MacPherson, P. S., Enns, M. W., & McWilliams, L. A. (2003). Neuroticism and self-criticism associated with posttraumatic stress disorder in a nationally representative sample. *Behaviour Research and Therapy*, *42*(1), 105–114. https://doi.org/10.1016/s0005-7967(03)00105-0

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage publications.

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Human Language Technologies, Volume 1*(4171–4186.). https://doi.org/10.18653/v1/n19-1423

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

Ferguson, L. J., Kowalski, K. C., Mack, D. E., & Sabiston, C. M. (2014). Exploring Self-Compassion and Eudaimonic Well-Being in Young Women Athletes. *Journal of Sport & Exercise Psychology*, *36*(2), 203–216. https://doi.org/10.1123/jsep.2013-0096

Finlay-Jones, A., Bluth, K., & Neff, K. (2023). Handbook of Self-Compassion. In *Mindfulness in behavioral health*. https://doi.org/10.1007/978-3-031-22348-8

Fok, C. C. T., & Henry, D. (2015). Increasing the sensitivity of measures to change. *Prevention Science*, *16*(7), 978–986. https://doi.org/10.1007/s11121-015-0545-z

Gerhardt, B. C., Serra, J. G., Zimmer, C., & Arteche, A. X. (2024). Role of self-criticism in postpartum mental health: a network analysis. *Psicologia Reflexão E Crítica*, *37*(1). https://doi.org/10.1186/s41155-024-00321-2

Germer, C. K., & Neff, K. D. (2013a). Self-Compassion in clinical practice. *Journal of Clinical Psychology*, *69*(8), 856–867. https://doi.org/10.1002/jclp.22021

Germer, C. K., & Neff, K. D. (2013b). Self-Compassion in clinical practice. *Journal of Clinical Psychology*, *69*(8), 856–867. https://doi.org/10.1002/jclp.22021

Gilbert, P. (2005). Compassion. In *Routledge eBooks*. https://doi.org/10.4324/9780203003459

Gorvankolla, N. a. K., & S, N. R. B. (2017). Application of text mining in Effective document analysis: Advantages, challenges, techniques and tools. *International Journal of Engineering Research And*, *V6*(04). https://doi.org/10.17577/ijertv6is040078

Grazzani, I., & Ornaghi, V. (2011). Emotional state talk and emotion understanding: a training study with preschool children. *Journal of Child Language*, *38*(5), 1124–1139. https://doi.org/10.1017/s0305000910000772

Hashimi, H., Hafez, A., & Mathkour, H. (2015). Selection criteria for text mining approaches. *Computers in Human Behavior*, *51*, 729–733. https://doi.org/10.1016/j.chb.2014.10.062

Have, M. T., Tuithof, M., Van Dorsselaer, S., Schouten, F., Luik, A. I., & De Graaf, R. (2023). Prevalence and trends of common mental disorders from 2007-2009 to 2019-2022: results from the Netherlands Mental Health Survey and Incidence Studies (NEMESIS), including comparison of prevalence rates before vs. during the COVID-19 pandemic. *World Psychiatry/World Psychiatry*, *22*(2), 275–285. https://doi.org/10.1002/wps.21087

Iliev, R., Dehghani, M., & Sagi, E. (2014a). Automated text analysis in psychology: methods, applications, and future developments. *Language and Cognition*, *7*(2), 265–290. https://doi.org/10.1017/langcog.2014.30

Iliev, R., Dehghani, M., & Sagi, E. (2014b). Automated text analysis in psychology: methods, applications, and future developments. *Language and Cognition*, *7*(2), 265–290. https://doi.org/10.1017/langcog.2014.30

Ji, Q., & Raney, A. A. (2020a). Developing and validating the self-transcendent emotion dictionary for text analysis. *PloS One*, *15*(9), e0239050. https://doi.org/10.1371/journal.pone.0239050

Ji, Q., & Raney, A. A. (2020b). Developing and validating the self-transcendent emotion dictionary for text analysis. *PloS One*, *15*(9), e0239050. https://doi.org/10.1371/journal.pone.0239050

Ji, Q., & Raney, A. A. (2020c). Developing and validating the self-transcendent emotion dictionary for text analysis. *PloS One*, *15*(9), e0239050. https://doi.org/10.1371/journal.pone.0239050

Ji, Q., & Raney, A. A. (2020d). Developing and validating the self-transcendent emotion dictionary for text analysis. *PLoS ONE*, *15*(9), e0239050. https://doi.org/10.1371/journal.pone.0239050

Jusoh, S. (2018). A study on NLP applications and ambiguity problems. *Journal of Theoretical & Applied Information Technology*, *96*(6).

Katz, L. F., & Gurtovenko, K. (2015). Posttraumatic stress and emotion regulation in survivors of intimate partner violence. *Journal of Family Psychology*, *29*(4), 528–536. https://doi.org/10.1037/fam0000128

Kazis, L. E., Anderson, J. J., & Meenan, R. F. (1989). Effect sizes for interpreting changes in health status. *Medical Care*, *27*(Supplement), S178–S189. https://doi.org/10.1097/00005650-198903001-00015

Kronthaler, F., & Zöllner, S. (2020). Data Analysis with RStudio. In *Springer eBooks*. https://doi.org/10.1007/978-3-662-62518-7

Kyriazos, T. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, *9*(2207–2230). https://doi.org/10.4236/psych.2018.98126.

Löw, C. A., Schauenburg, H., & Dinger, U. (2019). 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

Marshall, M. B., Zuroff, D. C., McBride, C., & Bagby, R. M. (2008). Self-criticism predicts differential response to treatment for major depression. *Journal of Clinical Psychology*, *64*(3), 231–244. https://doi.org/10.1002/jclp.20438

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A., De Schryver, M., De Winne, J., & Brysbaert, M. (2012). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior Research Methods*, *45*(1), 169–177. https://doi.org/10.3758/s13428-012-0243-8

Muris, P. (2015). A protective factor against mental health problems in youths? A critical note on the assessment of Self-Compassion. *Journal of Child and Family Studies*, *25*(5), 1461–1465. https://doi.org/10.1007/s10826-015-0315-3

Neff, K. D. (2011). Self-Compassion, Self-Esteem, and Well-Being. *Social and Personality Psychology Compass*, *5*(1), 1–12. https://doi.org/10.1111/j.1751-9004.2010.00330.x

Newton, P. E. (2012). Clarifying the consensus definition of validity. *Measurement Interdisciplinary Research and Perspectives*, *10*(1–2), 1–29. https://doi.org/10.1080/15366367.2012.669666

Pennebaker, J., Booth, R., & Francis, M. (2007). Linguistic Inquiry and Word Count (LIWC2007). *Pennebaker Conglomerates*. http://homepage.psy.utexas.edu/HomePage/Faculty/Pennebaker/Reprints/LIWC2007_OperatorManual.pdf

Ponizovskiy, V., Ardag, M., Grigoryan, L., Boyd, R., Dobewall, H., & Holtz, P. (2020). Development and Validation of the Personal Values Dictionary: A Theory–Driven tool for investigating references to basic human values in text. *European Journal of Personality*, *34*(5), 885–902. https://doi.org/10.1002/per.2294

Potter, R. F., Yap, A. K., Francis, A. J. P., & Schuster, S. (2014). Self-Compassion Mediates the Relationship between Parentalcriticism and Social Anxiety. *Revista Internacional De Psicología Y Terapia Psicológica/Revista Internacional De Psicología Y Terapia PsicolóGica*, *14*(1), 33–43. https://www.ijpsy.com/volumen14/num1/372/self-compassion-mediates-the-relationship-EN.pdf

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2010). Construction and factorial validation of a short form of the Self-Compassion Scale. *Clinical Psychology & Psychotherapy*, *18*(3), 250–255. https://doi.org/10.1002/cpp.702

Rips, J.M. (2016) *Self-Criticism and Self-Compassion: Two sides of the same coin?*

RStudio Team (2022). RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Salecha, A., Ireland, M. E., Subrahmanya, S., Sedoc, J., Ungar, L. H., & Eichstaedt, J. C. (2024). Large language models show human-like social desirability biases in survey responses. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2405.06058

Schuling, R., Huijbers, M. J., Van Ravesteijn, H., Donders, R., Kuyken, W., & Speckens, A. E. (2016). A parallel-group, randomized controlled trial into the effectiveness of Mindfulness-Based Compassionate Living (MBCL) compared to treatment-as-usual in recurrent depression: Trial design and protocol. *Contemporary Clinical Trials*, *50*, 77–83. https://doi.org/10.1016/j.cct.2016.07.014

Sevilla-Liu, A. (2022). Understanding Self-Compassion within Narrative Identity: The Struggles of Japanese Students with Measuring Up. *The Qualitative Report*. https://doi.org/10.46743/2160-3715/2022.5602

Strauss, C., Taylor, B. L., Gu, J., Kuyken, W., Baer, R., Jones, F., & Cavanagh, K. (2016). What is compassion and how can we measure it? A review of definitions and measures. *Clinical Psychology Review*, *47*, 15–27. https://doi.org/10.1016/j.cpr.2016.05.004

Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM SIGMOD Record*, *36*(3), 23–34. https://doi.org/10.1145/1324185.1324190

Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, *70*(9), 1970–1979. https://doi.org/10.1111/jan.12402

Suh, H., & Chong, S. S. (2021). What predicts meaning in life? the role of perfectionistic personality and Self-Compassion. *Journal of Constructivist Psychology*, *35*(2), 719–733. https://doi.org/10.1080/10720537.2020.1865854

Tausczik, Y. R., & Pennebaker, J. W. (2009). The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology*, *29*(1), 24–54. https://doi.org/10.1177/0261927x09351676

Thakur, K., & Kumar, V. (2020). An Overview of Text Mining: Application and Free Software Tools. *Library Waves*, *6*(2), 53–59. http://www.librarywaves.com/index.php/lw/article/view/90

Wakelin, K. E., Perman, G., & Simonds, L. M. (2021). Effectiveness of self-compassion-related interventions for reducing self-criticism: A systematic review and meta-analysis. *Clinical Psychology & Psychotherapy*, *29*(1), 1–25. https://doi.org/10.1002/cpp.2586

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology*, *84*(3), 608–618. https://doi.org/10.1037/0022-3514.84.3.608

Wickham H, François R, Henry L, Müller K, Vaughan D (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.2, <https://CRAN.R-project.org/package=dplyr>.

Wickham H, Miller E, Smith D (2023). _haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files_. R package version 2.5.4, <https://CRAN.R-project.org/package=haven>.

Wickham H, Vaughan D, Girlich M (2023). _tidyr: Tidy Messy Data_. R package version 1.3.0, <https://CRAN.R-project.org/package=tidyr>.

Wickham H, (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Zhang, H., Watson-Singleton, N. N., Pollard, S. E., Pittman, D. M., Lamis, D. A., Fischer, N. L., Patterson, B., & Kaslow, N. J. (2017). Self-Criticism and Depressive Symptoms: Mediating role of Self-Compassion. *Omega*, *80*(2), 202–223. https://doi.org/10.1177/0030222817729609

Zhang, J. W., Chen, S., & Shakur, T. K. T. (2019). From Me to You: Self-Compassion predicts acceptance of own and others' imperfections. *Personality and Social Psychology Bulletin*, *46*(2), 228–242. https://doi.org/10.1177/0146167219853846

Žižka, J., Dařena, F., & Svoboda, A. (2019). *Text Mining with Machine Learning : Principles and Techniques*. https://www.bookdepository.com/Text-Mining-with-Machine-Learning-Jan-Zizka/9781032086217

Appendix A

Synonyms for SC and Their Translations

Table A1

List of synonyms for self-compassion and their translations (self-kindness, mindfulness & common humanity)

Synonyms	Translation	Source
Compassion	Mededogen	Dutch psy. articles
Kindness	Vriendelijkheid	Van Dale
Self-kindness	Zelfvriendelijkheid	Dutch linguistic rules
Understanding	Begrip	Thesaurus
Self-understanding	zelfbegrip	Thesaurus
Care	zorg	Van Dale
Self-care	zelfzorg	Dutch linguistic rules
Gentleness	zachtheid	Van Dale
Self-gentleness	zelfzachtheid	Dutch linguistic rules
Forgiveness	vergeving	Van Dale
Self-forgiveness	zelfvergeving	Dutch linguistic rules
Support	ondersteuning	Thesaurus
Self-support	zelfondersteuning	Dutch linguistic rules
Patience	geduld	Van Dale
Self-patience	zelfgeduld	Dutch linguistic rules
Warmth	warmte	Van Dale
Self-Warmth	zelfwarmte	Dutch linguistic rules
Acceptance	acceptatie	Thesaurus
Self-acceptance	zelfacceptatie	Dutch linguistic rules
Non judgemental	Niet-oordelend	Dutch psy. articles

Comfort	troost	Van Dale
Self-comfort	zelftroost	Dutch linguistic rules
Mindfullness	mindfulness	Dutch psy. articles
Non-criticism	Niet-kritisch	Dutch linguistic rules
Encouragement	aanmoediging	Van Dale
Self-encouragement	zelfaanmoediging	Dutch linguistic rules
Resillience	veerkracht	Thesaurus
Self-resillience	zelfveerkracht	Dutch linguistic rules
Healing	genezing	Van Dale
Self-Healing	zelfgenezing	Dutch linguistic rules
Empathy	empathie	Dutch psy. articles
Self-empathy	zelfempathie	Dutch linguistic rules
Tolerance	tolerantie	Thesaurus
Self- tolerance	zelftolerantie	Dutch linguistic rules
“Kind to myself”	Vriendelijk voor mezelf	Dutch psy. articles
Patient	geduldig	Thesaurus
Shared Human Experience**	Gedeelde menselijke ervaring	Google translate, Dutch Forum (Taalunie)
Collective Human Experience**	Collectieve menselijke ervaring	DeepL translator & Dutch art. (Groene Amsterdammer)
Universal Human Condition**	Universele menselijke conditie	Google translate & Dutch lit.
Human Interconnectedness**	Menselijke verbondenheid	DeepL translator
Human Commonality**	Menselijke gemeenschappelijkheid	DeepL translator & Dutch academic lit.
Shared Vulnerability**	Gedeelde kwetsbaarheid	Google translate & Dutch psy. articles
Human Solidarity**	Menselijke solidariteit	Google translate & Dutch art. (Volkskrant)
Universal Suffering**	Universeel lijden	Dutch art (Boekblad) & google translate
Self-approval	zelfgoedkeuring	Dutch psy. articles

Self-love	zelfliefde	Van Dale
Love of self	Liefde voor zichzelf	Dutch linguistic rules
Love for oneself	Liefde voor jezelf	Thesaurus
Forgiving yourself	Jezelf vergeven	Thesaurus
Loving yourself	Jezelf liefhebben	Van Dale
Love yourself more	Meer van jezelf houden	Dutch linguistic rules
Good to yourself	Goed voor jezelf	Van Dale
Love of oneself	Liefde voor jezelf	Dutch linguistic rules
Care about yourself	Zorg voor jezelf	Van Dale
Selfcare	zelfzorg	Dutch linguistic rules
Care about myself	Zorg voor mezelf	Thesaurus
Self-gentleness	zelfzachtheid	Dutch linguistic rules
Ordinary humanity**	Gewone menselijkheid	Thesaurus
Everyday humanity**	Dagelijkse menselijkheid	Thesaurus
Shared humanity**	Gedeelde menselijkheid	Thesaurus
Common people**	Gewone mensen	Thesaurus
Mere humanity**	Lauter menselijkheid	Thesaurus
Our shared world**	Onze gedeelde wereld	Thesaurus
Mutual compassion**	Wederzijdse compassie	Thesaurus
Typical humanity**	Typische menselijk	Thesaurus
Common mercy**	Gewone bermhartigheid	Thesaurus
Accepted	Aanvaard	Thesaurus
Commonplace**	Alledaags	Thesaurus
Everyday**	Dagelijks	Thesaurus
Familiar**	Vertrouwd	Thesaurus

Frequent**	Frequent	Thesaurus
Natural**	Natuurlijk	Thesaurus
Prevailing**	Heersend	Thesaurus
Prevalent**	wijdverspreid	Thesaurus
Routine**	Routine	Thesaurus
Simple**	Eenvoudig	Thesaurus
Typical**	Typisch	Thesaurus
Universal**	Universeel	Thesaurus
"Treating yourself like you would treat a good friend."	Jezelf behandelen zoals je een goede vriend zou behandelen	Dutch article & google translate
"Being understanding and not too hard on yourself."	Begripvol zijn en niet te hard voor jezelf	Google translate & Dutch article
"Being okay with not being perfect."	Oke zijn met niet perfect zijn	DeepL translate & Dutch article
"Backing yourself up during tough times."	Jezelf ondersteunen tijdens moeilijke tijden	Google translate & Dutch article
"Being gentle and caring to yourself just like to loved ones."	Lief en zorgzaam voor jezelf zijn, net zoals voor geliefden	DeepL translator & Dutch article
"Calming yourself down when you're stressed."	Jezelf kalmeren als je gestresst bent	Google translate & Dutch article
"Taking care of yourself mentally and emotionally."	Zorgen voor jezelf, mentaal en emotioneel	DeepL translator & Dutch article
"We all go through similar struggles." **	We gaan allemaal door soortgelijke strijd	Google translate & Dutch article
“What we go through as humans.” **	Wat we doormaken als mensen	DeepL translate & Dutch article

"Being human comes with ups and downs."**	Mens-zijn brengt ups en downs met zich mee	Google translate & Dutch article
"We're all connected in some way." **	We zijn allemaal op de een of andere manier verbonden	DeepL translate & Dutch article
"We have more in common than we think."**	We hebben meer gemeen dan we denken	Google translate & Dutch article
"Everyone has moments of weakness." **	Iedereen heft momenten van zwakte	DeepL translate & Dutch article
"Standing together as humans."	Samen staan als mensen	DeepL translate & Dutch article
"Everyone faces hardships."	Iedereen wordt geconfronteerd met tegenslagen	Google translate & Dutch article
"Realizing everyone makes mistakes."	Beseffen dat iedereen fouten maakt	DeepL translate & Dutch article

Appendix B

Instruction Manual 'Judgement Phase'

Concept definition of Self-compassion (SC) by Neff, (2011)

"Kristen Neff, argued that compassion can also be turned towards the self and created a model for SC that consists of three facets: *mindfulness* versus overidentification, *self-kindness* versus self-judgement and *common humanity* versus isolation (Neff, 2011). These components work together to foster a self-compassionate mindset when faced with mistakes made by oneself, feelings of inadequacy, or different challenging situations in life (Neff, 2011)."

Concept definition of Self-criticism (SCr) By Potter et al. (2014)

As SC highly correlates with overall well-being (Neff, 2011) and SCr with depressive symptoms (Zhang et al., 2017), SCr can be viewed as the opposite end of SC. For this reason, we have included synonyms for this concept as well. Moreover, SCr entails A damaging type of self-evaluation that involves experiencing shame and emotions of worthlessness, feeling

like a failure, and sensing other people's rejection and judgment based on how well one performs in daily tasks (Potter et al., 2014).

Instruction judgement phase:

For this phase of the research, it is required that the lexicon created for the concepts of self-compassion and self-criticism is checked by 'experts' to gain intersubjectivity and reliability. The language central to this research is Dutch, meaning that the lexicon should be judged in Dutch (the English translation is added for the purpose of later publication). Below, there are four tables presented. Tables 1-3 each present synonyms for one separate facet of Self-compassion (Mindfulness, Self-kindness & Common humanity). The fourth table represents synonyms for self-criticism. Please decide per synonym whether you believe it belongs to the lexicon. Only check the box when you consider the synonym NOT to belong. Also, it is possible to write a remark. However, this is not mandatory. Finally, if you feel synonyms or concepts are still missing, please add them to the (fitting) table.

Table B1

Synonyms for the facet of self-compassion: self-kindness (or encompassing the whole concept)

Synonyms	**Expert remarks**	**Check box if <u>NO</u>**
Mededogen		☐
Vriendelijkheid		☐
Zelfvriendelijkheid		☐
Begrip		☐
Zelfbegrip		☐
Zorg		☐
Zelfzorg		☐
Zachtheid		☐
Zelfzachtheid		☐
Vergeving		☐
Zelfvergeving		☐

Ondersteuning	☐
Zelfondersteuning	☐
Geduld	☐
Zelfgeduld	☐
Warmte	☐
Zelfwarmte	☐
Troost	☐
Zelftroost	☐
Aanmoediging	☐
Zelfaanmoediging	☐
Veerkracht	☐
Zelfveerkracht	☐
Genezing	☐
Zelfgenezing	☐
Empathie	☐
Zelfempathie	☐
Tolerantie	☐
Zelftolerantie	☐
Vriendelijk voor mezelf	☐
Geduldig	☐
Zelfgoedkeuring	☐
Zelfliefde	☐
Liefde voor zichzelf	☐
Jezelf vergeven	☐
Jezelf liefhebben	☐
Meer van jezelf houden	☐
Goed voor jezelf	☐

Liefde voor jezelf	☐
Zorg voor jezelf	☐
Zelfzorg	☐
Zorg voor mezelf	☐
Zelfzachtheid	☐
…	
…	
…	
…	
…	
…	
…	
…	
…	
…	

Table B2

Synonyms (words and sentences) for the facet of self-compassion: mindfulness

Synonyms	**Expert remarks**	**Check box if NO**
Mindfulness		☐
Niet-oordelend		☐
Acceptatie		☐
zelfacceptatie		☐
Niet-kritisch		☐
"Begripvol zijn en niet te hard voor jezelf"		☐
"Oke zijn met niet perfect zijn"		☐

“Jezelf kalmeren als je gestresst bent”	☐
“Jezelf behandelen zoals je een goede vriend zou behandelen”	☐
“Jezelf ondersteunen tijdens moeilijke tijden”	☐
“Lief en zorgzaam voor jezelf zijn, net zoals voor geliefden”	☐
“Zorgen voor jezelf, mentaal en emotioneel”	☐
Begrijpen	☐
Tolereren	☐
Tolerant	☐
Omarmen	☐
Kalmeren	☐
Vergeven	☐
waarderen	☐
kracht	☐
Reflecteren	☐
…	
…	
…	
…	
…	

Table B3

Synonyms (words and sentences) for the facet of self-compassion: common humanity. Note. certain synonyms pertain to 'common' and some to 'humanity'., while others pertain to the combination of both.

Synonyms	**Expert remarks**	**Check box if <u>NO</u>**
Gewone menselijkheid		☐
Dagelijkse menselijkheid		☐
Gedeelde menselijkheid		☐
Gewone mensen		☐
Wederzijdse compassie		☐
Typische menselijk		☐
Gewone bermhartigheid		☐
Aanvaard		☐
Alledaags		☐
Dagelijks		☐
Vertrouwd		☐
Frequent		☐
Natuurlijk		☐
Heersend		☐
wijdverspreid		☐
Routine		☐
Eenvoudig		☐
Typisch		☐
Universeel		☐
Gedeelde menselijke ervaring		☐
Collectieve menselijke ervaring		☐

Universele menselijke conditie	☐
Menselijke verbondenheid	☐
Menselijke gemeenschappelijkheid	☐
Gedeelde kwetsbaarheid	☐
Menselijke solidariteit	☐
Universeel lijden	☐
"We gaan allemaal door soortgelijke strijd"	☐
"Wat we doormaken als mensen"	☐
"Mens-zijn brengt ups en downs met zich mee"	☐
"We zijn allemaal op de een of andere manier verbonden"	☐
"We hebben meer gemeen dan we denken"	☐
"Iedereen heft momenten van zwakte"	☐
"Samen staan als mensen"	☐
"Iedereen wordt geconfronteerd met tegenslagen"	☐
Delen	☐
Barmhartigheid	☐
Menselijk	☐
Menselijkheid	☐

Zorgzaam	☐
Gedeeld	☐
Verbondenheid	☐
Verbinding	☐
Solidariteit	☐
Gemeenschappelijk	☐
Iedereen	☐
Hulp	☐
Helpen	☐
…	
…	
…	

Table B4

Synonyms for self-criticism

Synonyms	**Expert remarks**	**Check box if <u>NO</u>**
kritiek		☐
Zelfkritiek		☐
Oordeel		☐
Hardheid		☐
Zelfhardheid		☐
Schuld		☐
Zelfschuld		☐
Verwaarlozing		☐
Zelfverwaarlozing		☐
Schaamte		☐

Zelfschade	☐
Schuldig	☐
Schuldig voelen	☐
Veroordeling	☐
Zelfveroordeling	☐
Straf	☐
Zelfstraf	☐
Twijfel	☐
Zelftwijfel	☐
Walging	☐
Zelfwalging	☐
Verwijt	☐
Zelfverwijt	☐
Zelfkritiek	☐
Onvergeeflijk	☐
Zelfonvergeeflijk	☐
Negativiteit	☐
Zelfnegativiteit	☐
Foutzoeken	☐
Zelffoutzoeken	☐
“bij jezelf de fout zoeken”	☐
“Gevoel van ontoereikendheid”	☐
Intolerant	☐
Ongedudlig	☐
“Gevoelens van falen”	☐

“Alleen voelen in falen”	☐
Obsederen	☐
Fixeren	☐
“Streng voor mezelf”	☐
Zelfverlagend	☐
Zelfonderzoek	☐
Zelfkritisch zijn	☐
Wantrouwen	☐
Aarzelend	☐
Afkeuring	☐
Zelfevaluatie	☐
Negatieve zelfevaluatie	☐
Zelfcommentaar	☐
Bezwaar	☐
Zelfaanval	☐
zelfbeschuldiging	☐
Beschuldinging	☐
Kritiek op kleinigheden	☐
…	
…	
…	
…	
…	
…	
…	
…	
…	
…	
…	

…

Table B5

Lexicon post judgement phase: SC (all facets) & SCr

Synonyms SC	Synonyms SCr
Mededogen	Oordeel
Vriendelijkheid	Oordelend*
Vriendelijk*	Kritisch*
Begrip*	Zelfkritisch*
Zelfbegrip*	Kritiek
Begripvol*	Zelfkritiek
Zorg	Lastigheden*
Zelfzorg	Lastigheid*
Zachtheid	Lastig*
Zacht*	Hardheid
Vergeving	Zelfhardheid
Zelfvergeving	Verharding*
Geduld	Verhard*
Warmte	Keihard*
Warm*	Hard*
Troost	Schuld
zelftroost	Zelfschuld
Aanmoediging	Schuldig
Veerkracht	Verwaarlozing

Kracht	Zelfverwaarlozing
Sterk*	Verwaarlozen*
Sterkte*	Verwaarloos*
Empathie	Schaamte
Empatisch*	Schaam*
Tolerantie	Veroordeling
Zelftolerantie	Vooroordeel*
Geduldig	Veroordeel*
Goedkeuring*	Straf
Zelfliefde	Straffend*
Eigenliefde*	Zelfstraf
Zelfzachtheid*	Gestraft*
Aardig*	Twijfel
Eigenwaarde*	Twijfelt*
Waarde*	Zelftwijfel
Waardes*	Walging
Waarden*	Zelfwalging
Waarderen	Walg*
Zelfwaarde	Walgen*
Zelfcompassie*	Walgt*
Compassie*	Verwijt
Compassievol*	Verweet*
Mildheid*	Zelfverwijt
Mild*	Onvergeeflijk
Liefde*	Negativiteit

Vergeven	Negatief*
Vergeef*	Zelfnegativiteit
Vergeeft*	Foutzoeken
Ondersteunen*	Fout*
Ondersteun*	Ontoereikendheid*
Ondersteuning	Ontoereikend*
Zelfondersteuning	Intolerant
Steun*	Ongeduldig
Omarmen	Ongeduld*
Omarm*	Falen*
Begrijpen	Faal*
Begrijp*	Faalt*
Tolereren	Faalde*
Tolereer*	Gefaald*
Tolerant	Beneden*
Genade*	Streng*
Mindfulness	Zelfverlagend
Mindful*	Verlagen*
Niet-oordelend	Verlaagd*
Acceptatie	Verlaag*
Zelfacceptatie	Zelfonderzoek
Accepteren*	Wantrouwen
Accepteer*	Afkeuring
Accepteert*	Afkeuren*
Geaccepteerd*	Keur*

Niet-kritisch	Zelfevaluatie
Kalmeren	Zelfcommentaar
Kalm*	Commentaar*
Kalmte*	Bezwaar
Lief*	Zelfaanval
Zorgzaam	Aanval*
Menselijkheid	Zelfbeschuldiging
Menselijk	Beschuldiging
Gedeeld	Beschuldig*
Delen	Perfect*
Barmhartigheid	Bekritiseren*
Barmhartig*	Bekritiseer*
Aanvaard	Bekritiserend*
Aanvaarden*	Bekritiseerd*
Alledaags	Zelfprojectie*
Dagelijks	Projecteren
Vertrouwd	Dom*
Natuurlijk	Onhandig*
Heersend	Stom*
Wijdverspreid	Idioot*
Typisch	Moeilijk*
Universeel	Moeilijkheid*
Verbondenheid	Moeilijkheden*
Verbonden*	Moeizaam*
Verbinding	Piekeren*

Kwetsbaarheid*	Pieker*
Kwetsbaar*	Piekeraar*
Solidariteit	Perfectionistisch*
Best*	Verwachting*
Goed*	Verwacht*
Houden*	Wegcijferen*
Doormaken*	Cijfer*
Meemaken*	Onzeker*
Ervaren*	Minderwaardig*
Gemeenschap*	Projecteer*
Gemeenschappelijkheid	Afkeur*
Gemeenschappelijk	Afgekeurd*
Iedereen	Aarzel*
Vertrouwen*	Aarzelend
Vertrouw*	Aarzelen*
Ademruimte*	Geaarzeld*
Ademen*	Verwijten*
Break*	Verwijt
Pauze*	Verweten*
Herstel*	Keuren*
Herstellen*	Sloom*
Terugkomen*	Slomer*
Mogen*	Traag*
Mag	Trager*
Open*	Lui*

Vooruit*	Boos*
Zingeving*	Gesloten*
Strijd*	Geslotenheid*
Gestreden*	Beoordelen*
Strijden*	Beoordeling*
Positiviteit*	Beoordeel*
Relativeren*	Koud*
Relativeer*	Kil*
Relativeert*	Strikt*
Gerelativeerd*	Precies*
Omgaan*	Nauwkeurig*
Harmonie	Nauwgezet*
Bewsutwording*	Consequent*
Bewustworden*	Scherp*
Bewustheid*	Onbuigzaam*
Aanmoedigen*	Veeleisend*
Liefdevol*	Moeilzaam*
Verdragen*	Kieskeurig*
Verdraag*	Onvrede*
Verdraagt*	Ontevreden*
Vrede*	Afwijzing*
Toelaten*	Afwijzen*
Toegelaten*	Afkraken*
Emoties*	Kraak*
Emotie*	Kraken*

Voelen*

Laten*

Toestaan*

Toegestaan*

Goedkeuren*

Gedogen*

Dogen*

Ondervinden*

Ondervindt*

Ondervind*

Aanvoelen*

Beseffen*

Besef*

Gewaarworden*

Gewaarwording*

Trotseren*

Trots*

Verduren*

Overgave*

Overgeven*

Verzorging*

Verzorg*

Verzorgen*

Aandacht*

Inzicht*

Inzichtelijk*
Rust*
Rustpunt*
Rusten*
Gerust*
Geruststellen*
Geruststelling*
Helpen*
Hulp*
Genoeg*
Reflecteren*
Gereflecteerd*
Reflecteer*
Reflecteert*
Reflectie*
Loslaten*
Los*
Pijn*
Geleden*
Lijden*
Leed*
Verzachten*

Appendix C

Percentages of the Constructs

Table C1.

Calculated percentages of the word count of the lexicon within each interview.

Participant	Pre_SC	Post_SC	Pre_SCr	Post_SCr
113	1.63%	6.85%	1.62%	0.68%
112	3.13%	4.63%	1.04%	0.78%
108	2.96%	0%	0%	1.46%
107	2.59%	4.63%	2.59%	0%
106	1.10%	2.03%	1.85%	2.32%
102	0%	3.76%	1.49%	0%
98	3.17%	3.34%	0.53%	0%
97	2.55%	0.52%	0%	0.52%
96	4.56%	2.06%	1.14%	0.25%
95	6.35%	3.53%	0.79%	0.88%
92	4.10%	2.67%	1.03%	0%
91	7.35%	2.98%	1.47%	0%
89	5.35%	1.46%	0.45%	0%
88	2.90%	2.02%	1.45%	0.34%
87	2.90%	2.55%	0%	1.91%
85	0.75%	0.33%	0.25%	0%
83	0.84%	1.50%	0.70%	1.50%
82	1.93%	1.48	1.21%	0.21%
81	1.66%	2.05%	0.66%	0.12%
80	2.34%	2.33%	0.88%	0%
79	1.06%	2.89%	0.75%	0.63%
75	10.00%	16.00%	0%	0%

73	2.10%	2.16%	0%	0%
41	9.26%	5.31%	1.85%	0%

Appendix D

Construct Validity: Post-Intervention Measures With SCS Data

First, Table D1 illustrates the correlations between the post-intervention scores of the lexical data and the post-intervention scores on the SCS. Additionally, the correlations with self-kindness (r = 0.008, p = 0.973) and common humanity (r = -0.033, p = 0.884) indicate weak associations between the notions. The link with mindfulness is also negative (r = -.0211, p = 0.346), suggesting minimal congruence with recognised SC indicators following the intervention. In addition, Figure D2 displays a scatterplot illustrating the relationship between post-intervention SC and the overall SCS scores following the intervention.

Table D1.

Correlations post-SC scores with post-SCS scores

Variable pair	**Correlation**	**P-value**
SC - Self-kindness	*.008*	*.973*
SC -_Common Humanity	*-.033*	*.884*
SC - Mindfulness	*-.211*	*.346*
SC - Self-judgement	*-.168*	*.455*
SC - Isolation	*.161*	*.474*
SC - Over-identification	*.039*	*.862*
SC - SCS_total	*-.055*	*.807*

Figure D2.

This exhibits the correlation analyses between the post-SC scores of the lexicon data and the total SCS data (T1)

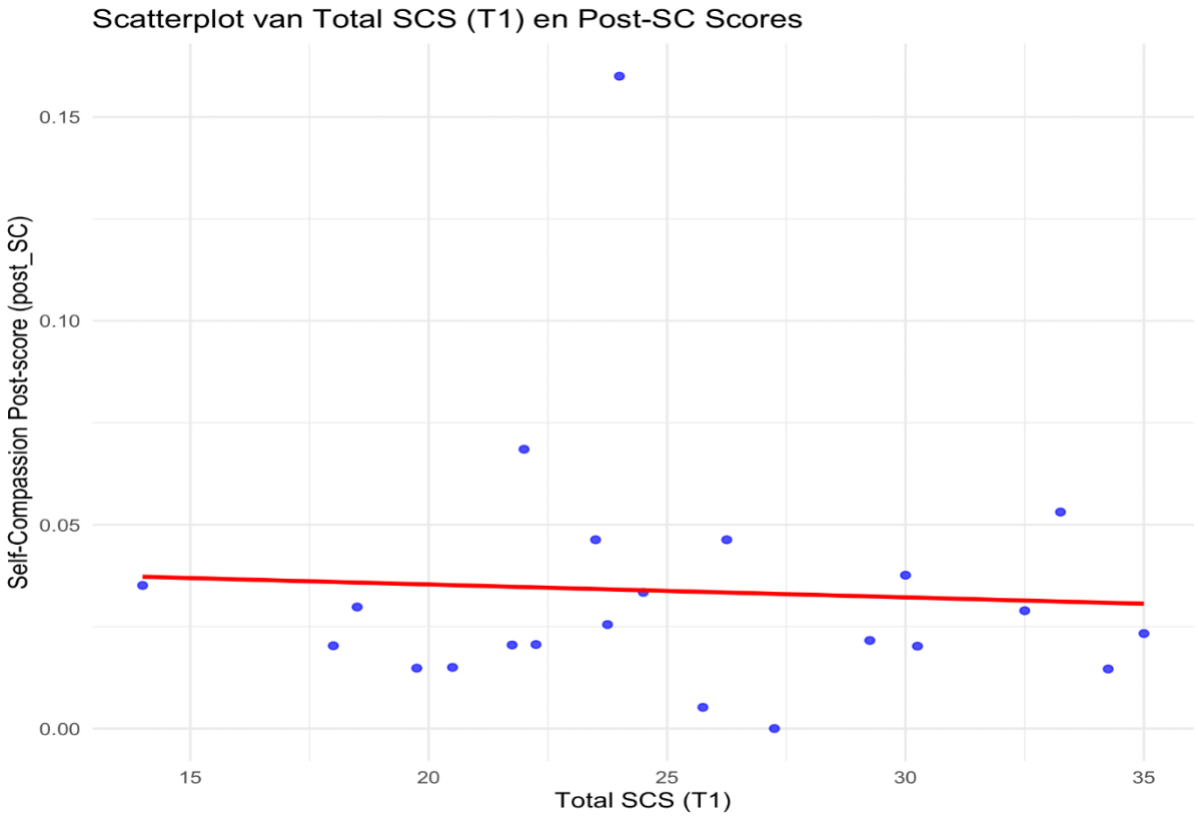


Table D3 illustrates the correlations between post-intervention SCr scores of the lexicon data and post-intervention SCS scores. The association with self-kindness is moderate and nearing significance ($r = -0.370$, $p = 0.090$), indicating a possible negative relationship, whereas the correlation with common humanity is modest and insignificant ($r = -0.177$, $p = 0.430$). Notably, the correlation with over-identification is significant ($r = -0.431$, $p = 0.045$), indicating that increased SCr is associated with lower tendencies to over-identify with negative emotions or aspects about one's storyline (Neff, 2011). Other correlations, such as mindfulness and common humanity, remain weak and insignificant. Furthermore, the

correlation between the total SCS scores at the post and the post-intervention SCr scores approaches significance, indicating that elevated SCr levels may correlate with diminished SC levels following the intervention (r = -.407, p = .060). Although the p-value was insignificant, the pattern suggests a potential association that warrants additional investigation.

Table D3.

Correlations between post-SCr scores and post-SCS scores

Variable pair	**Correlation**	**P-value**
SCr - Self-kindness	*-.370*	*.090*
SCr -_Common Humanity	*-.177*	*.430*
SCr - Mindfulness	*-.359*	*.101*
SCr - Self-judgement	*.230*	*.302*
SCr - Isolation	*.278*	*.210*
SCr - Over-identification	*.431*	*.045**
SCr - SCS_total	*-.407*	*.060*

Note. * = *significance,* *p < .05, **, p < .01, *** p < .001