**A systematic review of social connection inventories**

Bastien Paris

Independent Researcher

Debora Brickau

Ruhr University Bochum

Tetiana Stoianova

Université Grenoble Alpes

Maike Luhmann

Ruhr University Bochum

Christopher Mikton

World Health Organization

Julianne Holt-Lunstad

Brigham Young University

Marlies Maes

Utrecht University

Hans IJzerman

Annecy Behavioral Science Lab

**We wrote this registered report in the past tense to avoid errors when completing the Stage 2 Registered Report.**

**Author note:** Bastien Paris, Debora Brickau, and Tetyana Stoianova are shared first authors; authorship order was mutually agreed upon. Note that the methods and results were largely derived from the CORE Lab (2023); there is thus considerable overlap in these sections between that manuscript and ours.

**Abstract**

Social connection is vital to health and longevity. To date, a plethora of instruments exists to measure social connection, assessing a variety of aspects of social connection like loneliness, social isolation, or social support. For comparability and consistency of the published literature and for policy recommendations, consolidation and evaluation of the quality of measures is crucial. To answer the call for comparability, in Study 1a, we conducted a systematic review to create a database of social connection measures (*N*=xx) for its structure (*N*=xx), function (*N*=xx), and quality components (*N*=xx), spanning [YEAR] to [YEAR]; after which, in Study 1b, we assessed the heterogeneity of these existing measures through an item-content analysis relying both on human coders, as well as ChatGPT. We identified a total of XX item categories (XX for structure, XX for function, and XX for quality components) with a Jaccard index of XX for structure, XX for function, and XX for quality components. To answer the call for quality assessment, in Study 2a, we conducted a second systematic review on the measures found in Study 1a, creating a database documenting overall validity evidence. In Study 2b, we then evaluated the measurement properties using the COnsensus-based Standards for the Selection of health Measurement Instruments. We found the measurement properties to be [sufficient / insufficient / inconsistent / indeterminate], [sufficient / insufficient / inconsistent / indeterminate], and [sufficient / insufficient / inconsistent / indeterminate]; with [high/moderate/low/very low], [high/moderate/low/very low], and [high/moderate/low/very low] quality of evidence for the structure, function, and quality components, respectively. Finally, we identified the country of origin of the measures and the population groups with which they were developed, using data from Study 1a. Most of the measures were developed in [country name] (XX%) and for [add population characteristics] (XX%). [Overall conclusion].

**Keywords**: measurement, social connection, social isolation, loneliness, social support, systematic review, quality assessment

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Question** | **Hypothesis** | **Sampling Plan** | **Analysis Plan** | **Rationale for deciding the sensitivity of the test for confirming or disconfirming the hypothesis** | **Interpretation given different outcomes** | **Theory that could be shown wrong by the outcomes** |
| 1. To what extent do the items of the measures of social connection overlap within and between the different aspects of social connection (i.e., to what extent do the different measures assess the same or distinct constructs)? | We predict that measures of social connection will show – at best – weak overlap within and across two of the components (quality and function) of social connection, and therefore demonstrating, high heterogeneity.  We don’t have a strong hypothesis for the structure component. | We will conduct a systematic review to find measures for each of the three components of social connection based on a priori defined keywords. | We will estimate the item content overlap of measures via a Jaccard index correlation coefficient, within and across each component of social connection, thus for a total of four Jaccard index correlation coefficients.  We will estimate the item content overlap via a Jaccard index (0 = no overlap, 1 = full overlap). | We expect to capture most of the social connection measures available in the literature with our systematic review. | The potential outcomes are very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1. With any overlap lower than 0.60 *within* each category, we conclude that concept to measure mapping is of insufficient quality and we will conclude that a) better development-validation work is necessary and b) different psychometric approaches (e.g., network analyses) may be necessary. A low content overlap between measures and items further indicates heterogeneity across different measures of social connection (i.e., different measures assess different constructs) and an interchangeable use would pose a severe challenge.  For overlap 0.60 or higher, we conclude that concept to measure mapping is sufficient and the latent variable approach may be a right approach to understanding the concept.  For overlap no higher than 0.60 *across* categories, we will conclude that outcomes related to the three different components cannot be meta-analyzed jointly, but need to be separated in future meta-analyses.  Furthermore, a high content overlap between measures and items would indicate homogeneity across different measures of social connection (i.e., different measures assess the same construct) and would allow interchangeable use. A low content overlap between measures and items indicates heterogeneity across different measures of social connection (i.e., different measures assess different constructs) and an interchangeable use would pose a severe challenge. | If overlap is weak within categories, there is considerable chance for false positives in the existing literature in the absence of pre-registration. A low content overlap between measures and items indicates heterogeneity across different measures of social connection, implying different constructs being assessed. The interchangeable use of such measures poses a severe challenge. High heterogeneity suggests literature lacks comparability, potentially yielding idiosyncratic results when relying solely on a single measure. This poses a threat to the validity of research on social connection. |
| 1. How well validated are the measurement properties of social connection measures, and thus, what is the quality of social connection measurement? | We predict that measures of social connection will show insufficient evidence of great measurement properties. | We will rely on the articles data collected during the first systematic review and we will conduct an additional systematic review to find measurement properties evidence for each measure of social connection. | We will conduct a quality assessment of measurement properties using COSMIN methodology. We will 1) evaluate the quality of the methodology of studies assessing that measurement property, using the COSMIN risk of bias checklist; 2) evaluate the quality of the measures on that measurement property, following the COSMIN criteria for good measurement properties; 3) grade the quality of evidence, using the COSMIN-modified GRADE (Balshem et al., 2011) principles; 4) describe the interpretability and feasibility of the measures; 5) report the results following the PRISMA statement (Moher et al., 2015) and make recommendations on the measures based on these results. | We will rely on the COSMIN guidelines to judge measurement properties and quality of evidence. | The potential outcomes for rating each measurement property are sufficient, insufficient, inconsistent, and indeterminate. The potential outcomes for grading the quality of evidence are high, moderate, low, or very low. Following the COSMIN guidelines, a) we will recommend measures with at least sufficient rating of content validity and low-quality evidence grading of sufficient rating for internal consistency for use; b) we will not recommend measures with high-quality evidence grading of insufficient rating for any measurement property; and c) we will describe measures that don't fall in a) or b) as having potential for being recommended, but requires additional research. We will provide additional shortcomings in case they exist (e.g., lack of measurement invariance for global populations) | A low quality of measurement properties would suggest that portions of the literature are likely to be not reliable, which would pose a threat to replicability and generalizability. Conversely, a high measurement quality would indicate high reliability of the literature, providing more evidence. |
| 1. In which countries did the measures of social connection originate? | We predict that measures of social connection mostly originate from the US. | We relied on the data collected during the first systematic review. | We will provide descriptive statistics regarding where the measures originated. | Not applicable. | If a measure was not developed in a given country, no measurement invariance testing was conducted for the country of origin and the target country, *and* no further attention was given to the cognitive response processes within that country, we will recommend that the measure does not apply to the other country. We will not test for each country individually, but will only provide descriptive information for the reader to evaluate for the country of interest. | If measures are insufficiently developed and evaluated in different countries, we will conclude that applicability of the measures across countries is limited. |
| 1. With which populations in mind were the measures of social connection developed? (e.g., race, socioeconomic status, sexuality, et cetera) | No hypothesis. | We relied on the data collected during the first systematic review. | We will analyze the population characteristics. We will report where the measures were developed and whether age, race, gender, sexuality composition, religious affiliations, and socioeconomic indicators (income, education, employment status) were reported from the sample for which the measure was developed, and if so, what they were. | Not applicable. | If a measure was not developed in a given population group, no measurement invariance testing was conducted for the population group of origin and the target population group, and no further attention was given to the cognitive response processes within that population group, we will recommend that the measure does not apply to the other population group. We will not test for each population group individually, but will only provide descriptive information for the reader to evaluate for the population group of interest. | If measures are insufficiently developed and evaluated across different population groups, we will conclude that applicability of the measures across population groups is limited. |

**A systematic review of social connection inventories**

How people form and maintain healthy relationships is one of the most crucial questions in modern social science. A lack of social connection impacts health and longevity similarly to other clinical risk factors (Holt-Lunstad et al., 2010; Pantell et al., 2013), and other research suggest that living alone (versus living with others) is associated with a 32 % increased risk of early death (Holt-Lunstad et al., 2015). The US Surgeon General Vivek Murthy found the issue of loneliness so troubling, he dubbed it “a strategic priority” (Hagland, 2019). Despite the increasing efforts, such as a worldwide initiative to combat loneliness (the Global Initiative on Loneliness and Connection), the UK government dedicating a minister to social contact (UK Government, 2018), and near-immediate access to others through social media, trends remain concerning. For example, estimates suggest loneliness has either remained stable (Hawkley et al., 2019) or increased slightly (Buecker et al., 2021).

Perhaps one of the reasons that loneliness has not decreased may be that the monitoring of it is poorly understood. Loneliness, and its overarching concept of social connection is, after all, multi-dimensional with no clear consensus on a single definition of the concept. Further, different measurements of social connection do not correlate well with one another, suggesting that in fact they are not measuring the same construct across studies (Holt-Lunstad, 2018; see also Maes et al., 2022). The interchangeable use of instruments that measure very different indicators of social connection (and its link to related concepts, like health) could therefore pose a threat to the very understanding of social connection (see also Pomeroy et al, 2023, calling for such a consensus).

It is still an open theoretical question to what extent measures of social connection are comparable or not and how well each measure performs on standards of measurement quality. We, therefore, conducted a systematic review of measurements of social connection for non-clinical and adult populations. In Study 1a, we created a database of measures assessing social connection after which, in Study 1b, we assessed the heterogeneity of these measures through an analysis of item-content overlap of the measurement instruments within and between different aspects of social connection (Study 1b; CORE Lab, 2023; Fried, 2017a, b). In Study 2a, we conducted a second search to create a database of articles that provide overall validity evidence on the measures of social connection we found, after which, in Study 2b, we conducted a comprehensive quality assessment of the reported psychometric properties of these measures using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) to derive overall ratings of *sufficient*, *insufficient*, *inconsistent*, or *indeterminate* summarizing all psychometric properties. Finally, we described from which countries the measures originated and with which kinds of non-clinical and adult populations they were originally developed and validated.

**What is social connection and how is it measured?**

Social connection is the overarching term explicating how people connect to one another, how lonely they feel, and how socially isolated they are. Holt-Lunstad (2018) organized this general umbrella term of social connection into three different components of social relations: structure, function, and quality (see also Holt-Lunstad et al., 2017). The structural component is quantitative in nature. Measures of social connection’s structure assess how many social relationships a person has, how many groups the person is a part of, or how frequently someone engages in social contact (e.g., the Social Network Index; Cohen et al., 1997). The functional component captures the extent to which others can be relied upon to meet various needs and goals (e.g., social support, perceived partner responsiveness; Crasta et al., 2021; or loneliness, Russell, 1996). The qualitative component encompasses positive and negative perceptions of people’s relationships (e.g., measures of romantic relationship quality, like the Dyadic Adjustment Scale; Spanier, 1976).

How social connection is measured tends to vary considerably within and across the structural, functional, and qualitative components. Structural indicators have, for instance, been measured through questions as varied as marital status, network size and diversity, living arrangements (Parkerson & Gutman, 2000), volunteer status (Shmotkin et al., 2003), a composite index of whether participants were married, were living with another person(s), had frequent physical or phone contact with family members, and were home alone for less than 2 hours per day (Rodríguez-Artalejo et al., 2006). The measurement of functional indicators can similarly vary, with questions ranging from “If I were sick and needed someone (friend, family member, or acquaintance) to take me to the doctor, I would have trouble finding someone.” (Cohen & Hoberman, 1983), to “Today, [target person’s name] tried to see where I was coming from” (Ruan et al., 2020) to “How often do you feel that there is no one you can turn to?” (Russell, 1996) to “How easy is it to get practical help from neighbors if you should need it?” (Kocalevent et al., 2018). The measurement of qualitative indicators can similarly vary from those that are focused specifically on negative qualities such as conflict or strain (e.g., “My partner and I have frequent conflicts”;Zacchilli et al., 2016), those that assess satisfaction such as (e.g., “I feel satisfied with our relationship”;Rusbult et al., 1998), and those that get at the intimacy, closeness, or depth of a relationship or interaction (e.g., “I would feel comfortable telling things that I do not tell other people”;Hook et al., 2003).

For example, for a very specific subcomponent of the qualitative component of social connection, romantic relationship quality, the CORE Lab (2023) conducted an assessment of just how varied the measurements are. They found the overlap between questionnaires to be weak and the measures therefore highly heterogeneous. It should thus come as no surprise that the measurements across the components of social connection tend to correlate only to a limited degree, with, for instance, single-item measures of loneliness and multiple-item measures correlating between .27 and .60, and measures of social support and loneliness correlating between -.04 and -.51 (Gallup, 2022). Further, researchers typically recognize that social connection measures are multi-dimensional (Holt-Lunstad, 2018) and that there is a need for consensus and some process of consolidation of measures (Pomeroy et al, 2023). We therefore conducted a systematic review, in which we expect measures across but also within the functional and qualitative components of social connection to be highly heterogeneous. We don’t have a clear hypothesis for the structural component, as we expect it to mostly consist of single-item measures (e.g., whether someone is married or not).

**Challenges to the applicability of measures: Heterogeneity, internal structure validity, and intended target groups**

A high degree of heterogeneity and, thus, a lack of conceptual clarity is at the heart of many measurement problems (Flake & Fried, 2020; Fried & Flake, 2018). A lack of conceptual clarity is often revealed through a plethora of different scales trying to capture the same concept. For instance, the CORE Lab’s (2023) systematic review detected 26 different instruments in total that all measure romantic relationship quality. This variety of measurement instruments corresponds to the literature usually reporting exploratory evidence, providing only fuzzy conceptual definitions or often no definitions at all (see also Delatorre & Wagner, 2020). From the 26 different instruments, the CORE Lab (2023) coded 25 different item categories that were only weakly overlapping, ranging in content from affection and love, to agreement about proper conduct, to conflict resolution, to sexuality, to family life/parenting, to compatibility in attitudes/preferences (see also Fried, 2017a, b). More specifically pertaining to the domain of loneliness, two systematic reviews are already suggestive of its heterogeneity of social connection measurements, finding 18 different loneliness measurement instruments (Bugallo-Carrera et al., 2023; Maes et al., 2022).

If researchers do not sufficiently well find a consensus on the concepts they are measuring, it decreases the likelihood that across studies, the same concept is measured, potentially threatening the validity of the research (Fried, 2017a, 2017b; Visontay et al., 2019; Weidman et al., 2017). A low item-content overlap between scales may therefore lead to idiosyncratic results across studies and may threaten the measurement validity of a large number of studies. A high-content overlap when concepts are supposedly distinct can lead to empirical overlap of constructs that are conceptually distinct. Either of these cases could render comparisons between studies hard, if not impossible (Fried, 2017a). The first focus of this systematic review is therefore to provide an assessment of the (lack of) overlap between measures and items of those measures to help understand to what extent studies on social connection are, in fact, comparable.

A consensus on which concepts are measured is necessary, but not sufficient for construct validity. The 2014 Standards for Educational and Psychological Testing have outlined five sources of validity evidence based on (1) test content, (2) response processes, (3) internal structure (i.e., internal structure validity), (4) relations to other variables (i.e., convergent and discriminant validity), and (5) consequences of testing (i.e., predictive validity). All five sources of evidence contribute to the validation process and researchers should determine which ones are most appropriate based on their focus or application (American Educational Research Association, 2014).

Narrowing in specifically on 3) internal structure validity, when the CORE Lab (2023) assessed the structure validity evidence of romantic relationship quality, they concluded that authors did not systematically report internal structure validity evidence, suggesting that romantic relationship quality is often reported without strong internal structure validity evidence. In the present systematic review, we will provide a broader evaluation, assessing the instruments according to the COSMIN guideline of patient-reported outcome measures (PROMs; Prinsen et al., 2018), evaluating content validity, methodological quality, instrument quality, instrument interpretability, and instrument feasibility, after which we will provide recommendations regarding the use of the instruments. This evaluation is the second focus of this systematic review, and we expected social connection measures to mostly show insufficient evidence of great measurement properties, largely in line with the results obtained by CORE Lab (2023).

Furthermore, it is widely known that psychology has a “generalizability” problem (Henrich et al., 2010), with editors and authors mostly in the United States and awards mostly going to researchers from the United States (IJzerman et al., 2021). US participants are often being considered the “default humans”, with researchers habitually (and inappropriately) generalizing from US participants to the world (Cheon et al., 2020). The incentives towards US dominance probably directly contributes to the rest of the world being understudied (Adetula et al., 2022; Silan et al., 2021). Both Delatorre and Wagner (2020) and the CORE Lab (2023) similarly observe that instruments assessing marital and relationship quality mostly originate in the United States. The approach of “generalizing almost exclusively from North America [and, to some extent, Europe] risks creating “unknown unknowns”, gaps that American and European psychologists inadvertently ignore because they do not consider the phenomena and processes that are outside their personal experience”; Adetula et al., (2022). Relatedly, Taylor et al. (2023) recently pointed out that research on loneliness and social isolation—two subcomponents of social connection—has been conducted unevenly across world regions and among different populations (e.g., ethnic groups, immigrant communities, or cultural groups). It is probable that similar problems are present in the broader field of social connection, which extends beyond psychology.

Take, for instance, a widely adopted theory on social connection, attachment theory. Traditionally, attachment theory focuses on parents, who are typically the primary caretakers of infants in higher-income countries. Among traditional families in Madagascar, however, infants interact almost exclusively with peer groups of older children, who end up being infant caretakers (Keller, 2018; Scheidecker, 2017). Attachment theory may have only very limited applicability in certain regions, and this also applies to the measures developed to test it. More generally, strong homogeneity in world regions that developed social connection measurements and in the population groups with which they were developed may be indicative of a miscalibration of those measurements to assess social connection across various human populations. A third focus of this systematic review is thus to identify where measures originated, and, relatedly, a fourth focus is with which kinds of population groups in mind (i.e., age, race, gender, sexuality, religious affiliation, and socioeconomic indicators) the measures were developed. In line with previous research, we expected social connection measures to mostly originate from the US (CORE Lab, 2023; Delatorre and Wagner, 2020; Henrich et al., 2010).

**Research Overview**

In the present article, we conducted two related systematic reviews to assess the quality and focus of social connection measurements through four steps: First in Study 1a, we created a database of measures to assess levels of social connection through a systematic review of instruments measuring at least one component (structure, function, or quality) of social connection. In Study 1b, we then examined the item-overlap between instruments, within and across each of the three different components of social connection. In Study 2a, we conducted a second systematic review, searching for measurement property evidence of the measures found in Study 1a. Finally, in Study 2b, we evaluated the evidence found using the COSMIN guidelines and identified in which country the measures originated and with which non-clinical population groups in mind they were developed (i.e., age, race, gender, sexuality, religious affiliation, and socioeconomic indicators).

**Study 1a**

**Methods**

In Study 1a, we created a database of social connection measures across the three different components of social connection: structure, function, and quality. We first conducted a literature review for each component (i.e., three in total) of social connection to select measures and supplemented this search by measures cited in previously published reviews (Bugallo-Carrera et al., 2023; CORE Lab, 2023; Holt-Lunstad et al., 2010; Maes et al., 2022; and Valtorta et al., 2016).

We selected self-reported measures based on pre-defined selection criteria. To ensure the reproducibility of our literature searches, we followed the recommendations provided by Maggio et al. (2011): We recorded 1) the databases on which we conducted the searches, 2) our selected search terms, 3) the dates on which we conducted our searches, 4) the full search strings with Boolean operators we used, and 5) the limits on the extraction of search results we set. The present systematic review aimed to be reproducible and fully transparent. In order to accomplish this goal, we made analysis scripts and the data for this study publicly available on our project page on the OSF (<https://osf.io/esxqm/>).

**Search strategies for finding development-validation articles for the structural, functional, and qualitative components of social connection**

The first systematic review is split into three separate searches to select measures assessing social connection. To guide our search terms, we developed working-definitions of the three different components of social connection from previous research (Holt-Lunstad, 2018, 2021; Holt-Lunstad et al., 2017):

1. Structural indicators of social connection: The connection to others via the existence of relationships and their roles (e.g., marital status, family size).
2. Functional indicators of social connection: A sense of connection that results from resources and functions provided or perceived to be available by social relationships (e.g., perceived social support, loneliness).
3. Qualitative indicators of social connection: The sense of connection to others that is based on positive and negative affective qualities (e.g., relationship satisfaction, intimacy, or conflict).[[1]](#footnote-2)

We recognize that these definitions are overlapping to some extent. We expect the categories and their respective definitions to be updated based on our systematic review. We coupled the social connection search terms with measurement search terms (e.g., “assessment”, “questionnaire”) and scale development/validation search terms (e.g., “confirmatory factor analysis”, “internal consistency”). For the structural component, we expected that many measures might be single-item measures, which are not prone to undergoing any work on development or validation (for instance, one is married or not; one is not married 5.4 points on a 5-item 7-point Likert scale). This led us to remove the scale development/validation search terms for that specific search[[2]](#footnote-3). Because of the omission of these search terms, our trial search for the structural component delivered more than 400,000 (often redundant) results across the three search databases, which was too large to review.

To ensure the review on structural indicators was feasible to conduct, we extracted a subset of the full search results totaling the average results detected for the functional and qualitative indicators (*N* = XX). We sorted the search results by relevance to reduce noise in the results and extracted search results stratifying for publication year to ensure a fair representation of articles across time. In order to do so, we 1) performed the structural search, 2) retrieved the total number of search results and the number of search results for each publication year, 3) computed the percentage of search results for each publication year, 4) computed the number of search results to retrieve for each publication year. Scripts and spreadsheets necessary to replicate this process are documented on our OSF Page: <https://osf.io/stmdb/>.[[3]](#footnote-4)

The finalized search strings can be found in Appendix A; we relied on the English, Dutch, Ukrainian, Russian, German, Turkish, Arabic, Spanish, Italian, and French language literatures[[4]](#footnote-5) using the following databases: Scopus, PubMed, and ProQuest (for ProQuest, we selected the Psychology Database, Social Science Database, ProQuest Dissertations, and Theses Global, Australia & New Zealand Database‎, Continental Europe Database‎, East & South Asia Database‎, East Europe, Central Europe Database‎, India Database‎, Latin America & Iberia Database‎, ‎Middle East & Africa Database‎‎, Turkey Database‎, and UK & Ireland Database‎).

**Selection criteria[[5]](#footnote-6)**

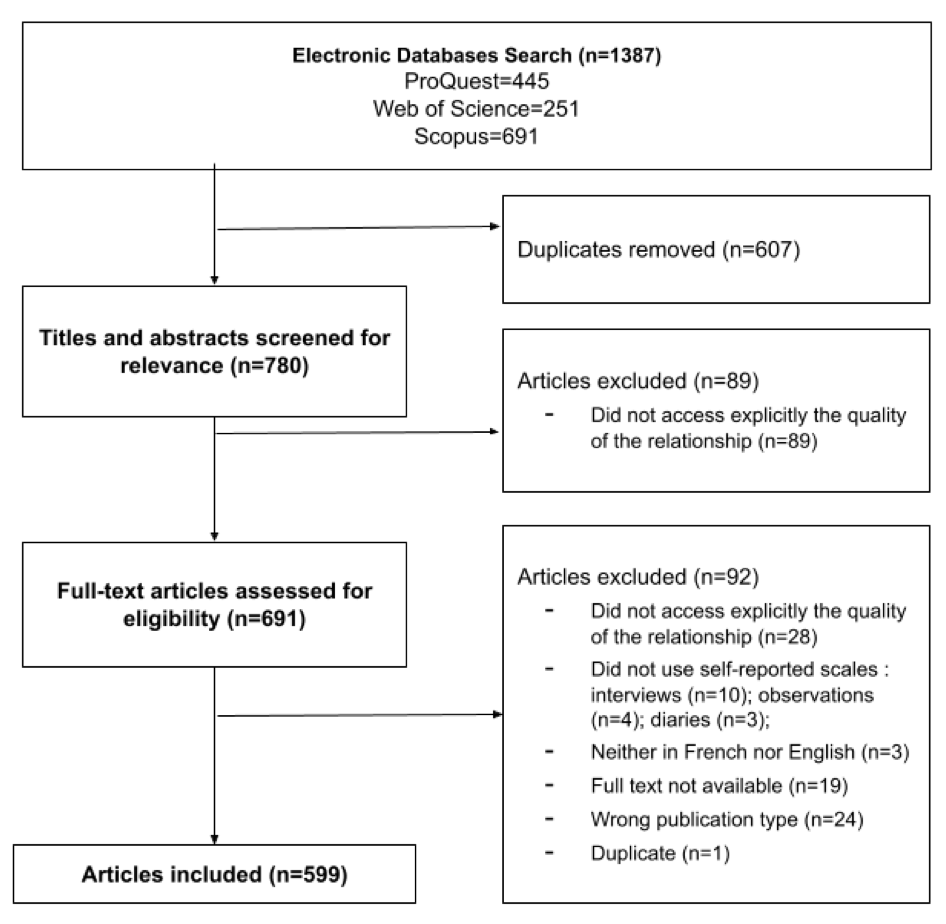
We included measures if 1) the original authors of the measure intended to measure social connection; 2) they were focused on adults (i.e., 18+); 3) the populations on which they were tested were non-clinical populations; 4) they were self-report measures that target people’s individual attitudes[[6]](#footnote-7); 5) the article of the measure was in English, Dutch, Ukrainian, Russian, German, Turkish, Arabic, Spanish, Italian, or French (the measure could be in any language); 6) for articles retrieved with the search targeting either the functional or qualitative indicators of social relationships, one goal of the paper in which the measure was presented was to develop/translate/validate the measure; 7) the full text was available; 8) the full questionnaire was available. We 9) have no limitations regarding the year of publication. The selection criteria can be found in Appendix B.

**Results**

For the three components of social connection together, we identified XX articles and first removed duplicates (*N* = XX; for a separation into the three categories of social connection, see Figures 1a-c), after which we screened titles and abstracts for relevance using Rayyan QCRI (Ouzzani et al., 2016), leading to a further removal of XX articles. Then, after reading the full text articles, we removed an additional XX articles. In the end, XX articles were initially included as eligible for further review.

Once the authors [ADD AUTHOR INITIALS ONCE COMPETED] completed the selection of the first 20%, two other authors each double-checked 10% of the procedure. Agreements with the first authors, assessed with Cohen’s κ (Landis & Koch, 1977), were [not satisfactory/satisfactory], reaching XX and XX. We resolved disagreements through discussion. We repeated this process once (thus having double-checked 40% in total). The initial XX articles broadly covered the structural, functional, and qualitative components of social connection. The agreements for the second round of selection were [not satisfactory/satisfactory], reaching XX and XX. [We will continue double-checking until an agreement of at least .80 is reached].

Articles were excluded based on the following reasons: 1) the original author(s) did not explicitly state that their instrument was intended to measure a component of social connection (*N* = XX), 2) the research was not focused on adults (*N* = XX), 3) the population tested was clinical (*N* = XX), 4) the measures were not self-report that target people’s individual attitudes (*N* = XX), 5) the article was not in English, Dutch, Ukrainian, Russian, German, Turkish, Arabic, Spanish, Italian, or French (*N* = XX), 6) for articles retrieved with the search targeting either the functional or qualitative components of social relationships, one goal of the paper was not to develop/translate/validate the measure (*N* = XX), 7) the full text was not available (*N* = XX), 8) the full questionnaire was not available (*N* = XX). The selection list (and their specific reasons for exclusion) of the articles can be found at our OSF Page: <https://osf.io/jsybf/>. The full selection process is described in Figure 1a-c.



*Figure 1.* Placeholder for our flowchart for the to-be-presented Figures 1a-c. This figure is from the CORE Lab (2023). We will insert three figures, one for each component of social connection (structure, function, and quality).

The measure selection process resulted in XX measures for the structural component of social connection, XX measures for the functional component of social connection, and XX measures for the qualitative component of social connection. In Tables 1a-c we summarized the measures characteristics for each of the indicators of social connection. Notably, we reported how many items the measures consisted of, whether the measures were unidimensional or multi-dimensional with their number of dimensions, and whether the authors recommended procedures to aggregate scores across items (e.g., sum score, mean score).

**Table 1a.** *Summary of Measures’ Characteristics Assessing Structural Indicators of Social Connection*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure | Authors | Publication Year | #Items | Uni/Multi? | #Dimensions | Sum Score? |
| XX | XX | XX | XX | XX | XX | XX |

*Table Note*.

**Table 1b.** *Summary of Measures’ Characteristics Assessing Functional Indicators of Social Connection*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure | Authors | Publication Year | #Items | Uni/Multi? | #Dimensions | Sum Score? |
| XX | XX | XX | XX | XX | XX | XX |

*Table Note*.

**Table 1c.** *Summary of Measures’ Characteristics Assessing Qualitative Indicators of Social Connection*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Measure | Authors | Publication Year | #Items | Uni/Multi? | #Dimensions | Sum Score? |
| XX | XX | XX | XX | XX | XX | XX |

*Table Note*.[[7]](#footnote-8)

[GIVEN THAT STUDY 1A WILL CONSIST OF THREE LITERATURE REVIEWS WITH NO ANALYSIS OF THE LITERATURE, NO DISCUSSION WILL BE ADDED BETWEEN STUDIES 1B AND 1B]

##### **Study 1b Item content analysis: Approach**

After selecting the measures, we estimated the item content overlap within *and* across each indicator of social connection, thus for a total of four analyses. The goal of the thematic content analysis was to assess to which degree items assess the same construct. We adopted this approach from Fried (2017a), who determined depression symptom overlap among seven depression scales via a Jaccard index (0 = no overlap, 1 = full overlap; for details see below), and from the CORE Lab (2023), who assessed item-content overlap in 26 scales that assess romantic relationship quality.

**Methods**

We first categorized the items into the different components of social connection. We first extracted all the XX items from the XX measures into a spreadsheet. Although we selected measures based on whether they measured structural, functional, or qualitative indicators of social connection, we then, item-by-item, judged which of the three components they belonged to (an item could belong to multiple components or to no component at all). For example, although a measure could *primarily* focus on the qualitative component of social connection, a few items may belong to the other components of social connection.

Once we had done the initial categorization into the three components of social connection of each item, we then categorized the items into more specific categories. We created our categories through an abductive logic. Authors [ADD INITIALS OF CODERS ONCE COMPLETED] carried out this process by creating categories based on the items and how they were worded. For example, (a) the items [“ITEM CONTENT”] [(MEASURE NAME)] and [“ITEM CONTENT”] [(MEASURE NAME)] were both listed under the category labeled as [“CATEGORY LABEL”]; whereas (b) the items [“ITEM CONTENT”] [(MEASURE NAME)] and [“ITEM CONTENT”] [(MEASURE NAME)] were both listed under the category labeled as [“CATEGORY LABEL”].

We deviated from our coding strategy from earlier work (CORE Lab, 2023), where coding into multiple categories was permitted. Even though this coding strategy may make some sense conceptually, in our experience, it is very difficult to have a precise measure of overlap, as it is very challenging to get interrater agreement higher than .60. In case the item could be classified into multiple categories, we either refined the category, or classified the item into a category that seemed to provide the best fit. We suspect this adjustment may somewhat inflate the final number of categories, potentially leading to an even weaker overlap between measures.

To ensure that the categories were accurate representations of the items, each of three other authors (XX, XX, and XX) independently cross-checked a separate 10% of the coding list, after which we calculated interrater agreement using Cohen’s κ (Landis & Koch, 1977). We repeated this process until we cross-checked at least 60% of the coding list and until the average agreement was at least .80. If disagreements arose, we either resolved the disagreement, changed the category description, or added new categories. This cross-checking process ensured that both the creation of the codebook and the categorization of items had collective agreement. After the final cross-check, average agreement was [not satisfactory/satisfactory] (κ = XX). [We continue this process until at least .80 agreement is reached, after which the main coder codes the rest. At least 60% of the items will be cross-checked].

As we expected heterogeneity among measures, we could not rule out the possibility that our ratings have been biased toward finding greater heterogeneity. To address that concern to some extent and to further strengthen the cross-checking process, we leveraged artificial intelligence to categorize the items in the categories we created. As the artificial intelligence would be blind to our hypotheses, a great interrater agreement with the artificial intelligence would suggest that our categorization judgments are both bias-free and accurate. Using Python scripts, we prompted gpt-3.5 (OpenAI, n.d.) with the categories we created and the items to categorize (see Appendix C for a template of the prompts we sent). To the best of our knowledge, no research formally assessed the performance of gpt-3.5 on this specific task. As such, results presented below should be taken with caution, though a test of gpt-3.5 performance using data from CORE Lab (2023) suggests that it is well suited to perform this task (κ = .80, see <https://osf.io/stmdb/> for scripts and data). After the final cross-check, average agreement with gpt-3.5 were [not satisfactory/satisfactory], reaching XX. These results suggest that our coding process is [biased/unbiased] and [reliable/unreliable]. A table displaying the agreement obtained with each coder can be found in Table 2.

Table 2. *Coding Agreements (Cohen’s* κ*) Obtained During the Coding Process for Specific Percentages of the Coding List. Bold values have been obtained after refining categories and resolving disagreements through discussion.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Coding list | Coder B | Coder C | Coder D | Coders’ average | GPT-3.5 |
| 0-10% | xx; **yy** |  |  |  | xx; **yy** |
| 10-20% |  | xx; **yy** |  |  | xx; **yy** |
| 20-30% |  |  | xx; **yy** |  | xx; **yy** |
| total (0-30%) |  |  |  | xx; **yy** | xx; **yy** |
| 30-40% | xx; **yy** |  |  |  | xx; **yy** |
| 40-50% |  | xx; **yy** |  |  | xx; **yy** |
| 50-60% |  |  | xx; **yy** |  | xx; **yy** |
| total (30-60%) |  |  |  | xx; **yy** | xx; **yy** |
| total (0-60%) |  |  |  | **yy** | **yy** |
| total (0-100%) |  |  |  |  | **yy** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

**Results**

**Data Analysis**

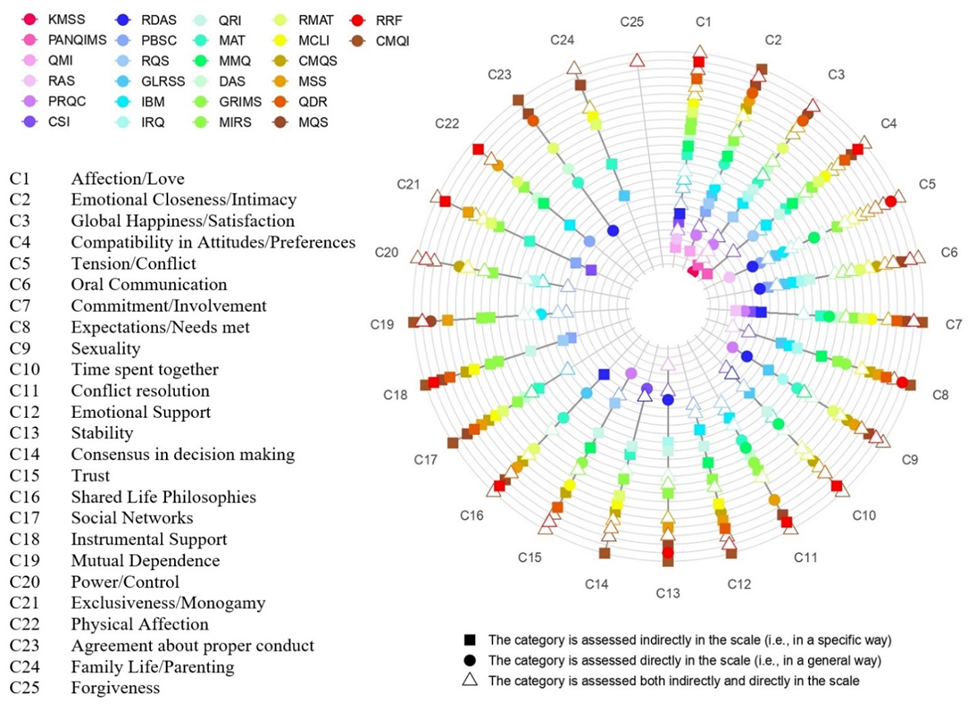
We used the Jaccard Index to estimate the content overlap a) within each component of social connection and b) across all three components of social connection (Fried, 2017a; the CORE Lab, 2023). This metric is a similarity coefficient specifically for binary data that ranges from 0 (no overlap among measures) to 1 (complete overlap). As described in detail in Fried (2017a), we calculated this coefficient with the formula *s/( u1 + u2 + s)*, where *s* is the number of categories two measures share, and *u1* and *u2* the number of categories that are unique to each of the two measures. We used the criteria of Evans (1996) for the Jaccard Index correlation coefficient: very weak 0.00–0.19, weak 0.20–0.39, moderate 0.40–0.59, strong 0.60–0.79, and very strong 0.80–1. In addition, we also calculated the rate of idiosyncratic categories per measure (i.e., categories that appear on no other measure). These analyses were conducted in R (Version X.X.X; R Core Team, 2022) and are available at our OSF Page: <https://osf.io/jsybf/>.

**Item categories**

The categorization process of all XX items from XX measures resulted in XX categories for the structural components of social connection, XX categories for the functional components of social connection, and XX categories for the qualitative components of social connection, for a total of XX categories. The measures covered a [narrow/medium/wide] range of content such as [Add example topics] (see Figure 2a-d). XX items were categorized as miscellaneous and thus were not included in the categories [Example item to be added]. Among the XX categories, [Category name 1] was the most common category across the measures (featured in XX measures), followed by [Category name 2] and [Category name 3]. Among the XX categories, [Category name 4] was the most common category for the structural component, [Category Name 5] was the most common for the functional component, and [Category Name 6] was the most common for the qualitative component. [To add if it applies: However, no measure captured all categories simultaneously, nor did one of the XX categories 3a-c.

**Item content overlap**

across the three components of social connection. We present overlap among all individual measures and mean overlap of each measure with all other XX measures in Tables 4a-c. [We will summarize points that may be worth highlighting from Tables 4a-c].



*Figure 2.* Placeholder for our Jaccard Index Figure for the to-be-presented Figures 2a-d. This figure is from the CORE Lab (2023). We will insert four figures, one for each component of social connection (structure, function, and quality) and one for an analysis across all three components. [Figure note to be added: Co-occurrence of XX thematic categories across XX measures assessing [structure, function, quality, or all three] components of social connection. Squares for a category indicate that a measure only assesses the category indirectly, circles indicate that a measure only assesses the category directly, and triangles indicate that a measure assesses the category both indirectly and directly. Shapes (circles, squares, and triangles) represent measures (from outer to inner shapes, with the number of categories assessed): [MEASURE\_A (n\_A), MEASUREE\_B (n\_B), …, MEASURE\_N ( n\_N)].

**Table 3a**

*Occurrence of categories and their number of items for the structure component of social connection*

|  |  |  |
| --- | --- | --- |
| Category | Appears in N measures | Composed of N items |
| XX | XX | XX |

**Table 3b**

*Occurrence of categories and their number of items for the function component of social connection*

|  |  |  |
| --- | --- | --- |
| Category | Appears in N measures | Composed of N items |
| XX | XX | XX |

**Table 3c**

*Occurrence of categories and their number of items for the quality component of social connection*

|  |  |  |
| --- | --- | --- |
| Category | Appears in N measures | Composed of N items |
| XX | XX | XX |

**Table 4a.** *Overlap of item content of XX measures assessing structural indicators of social connection*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure Name | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | … |
|  | 0.52 | XX | XX | XX | XX | XX | XX | XX | XX |  |
| Mean |  | XX | XX | XX | XX | XX | XX | XX | XX |  |

**Table 4b.** *Overlap of item content of XX measures assessing functional indicators of social connection*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure Name | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | … |
|  | 0.52 | XX | XX | XX | XX | XX | XX | XX | XX |  |
| Mean |  | XX | XX | XX | XX | XX | XX | XX | XX |  |

**Table 4c.** *Overlap of item content of XX measures assessing qualitative indicators of social connection*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure Name | 1 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | … |
|  | 0.52 | XX | XX | XX | XX | XX | XX | XX | XX |  |
| Mean |  | XX | XX | XX | XX | XX | XX | XX | XX |  |

**Discussion Study 1b**

[Discussion will be added following the analyses]

**Studies 2a and b**

In Studies 2a and b, we a) conducted a second systematic review of all measures detected in Study 1a to create a database for overall validity evidence, after which we b) provided a review on the measurement properties of the social connection measures included in Study 1a using the COSMIN guidelines of PROMs (Prinsen et al., 2018), provided an overview in which country the measures were developed, determined whether sample information was provided and if so, determined which populations the measures were developed with.

**Study 2a - Systematic review to evaluate the measurement properties of social connection measures**

**Methods**

**Literature search**

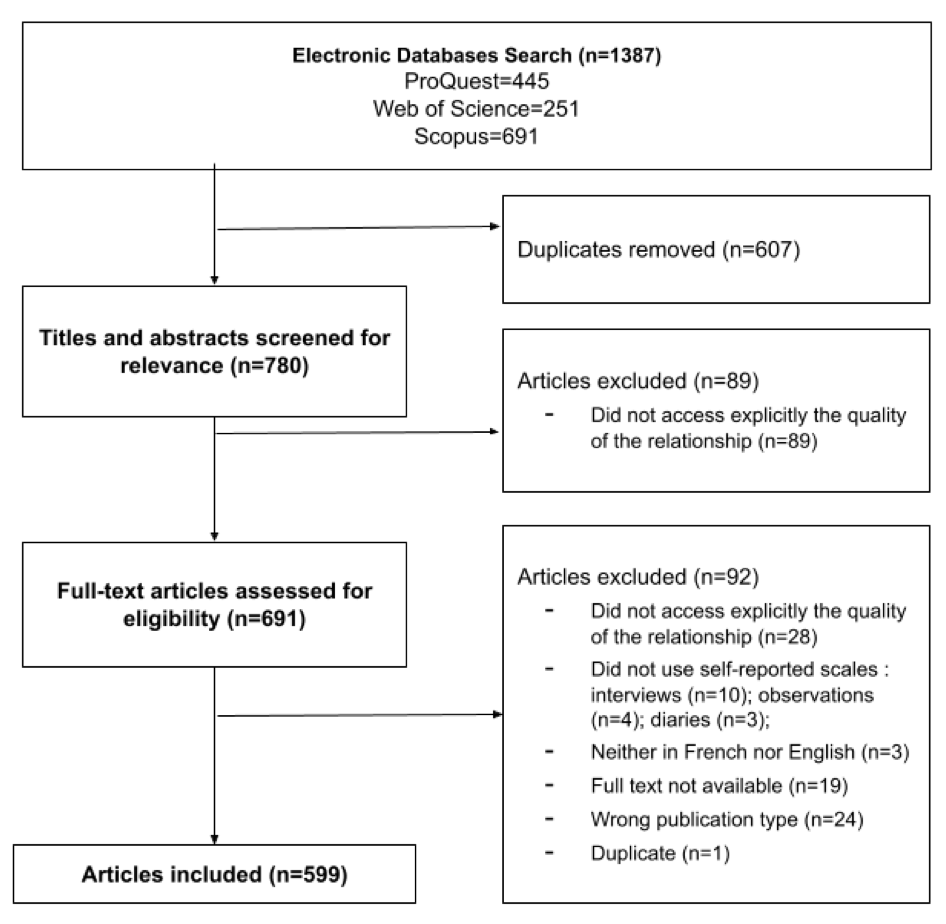
In Study 2a, we conducted a literature review based on Study 1a. Again, to ensure the reproducibility of our literature searches, we followed the recommendations provided by Maggio et al. (2011): We recorded 1) the databases on which we conducted the searches, 2) the search terms we selected, 3) the dates we conducted the searches, 4) the full search strings with Boolean operators we used, and 5) the limits on the extraction of search results we set. The present systematic review aimed to be reproducible and fully transparent. In order to accomplish this goal, we reported the review following the PRISMA statement (Moher et al., 2015). Analysis scripts and the data for this study are publicly available on our project page on the OSF (<https://osf.io/wfers/>)[[8]](#footnote-9).

We thus relied on the articles included in Study 1a and conducted an additional literature search for each measure, using a search string containing the name of the measure coupled with the measurement properties filters and the exclusion filters suggested in the COSMIN guidelines (Prinsen et al., 2018; Terwee et al., 2009). The definitive search strings can be found in Appendix D. We included any article that was specifically aimed at assessing at least one of the measurement properties included in the COSMIN taxonomy of measurement properties (Mokkink et al., 2010): Content validity, structural validity, internal consistency, cross-cultural validity, measurement invariance, reliability, measurement error, criterion validity, hypothesis testing for construct validity, and responsiveness. We selected articles based on a screening of title, abstract, and full-text (when appropriate) conducted by [rater initials will be added upon completion].

**Results**

**Results of the articles selection**

After having performed an additional search for each of the XX measures of social connection found in Study 1, we identified XX articles and first removed duplicates (*N* = XX), after which we screened titles and abstracts for relevance using Rayyan QCRI (Ouzzani et al., 2016), leading to a further removal of XX articles. Then, after reading the full text articles, we removed an additional XX articles. Once the first authors completed the selection of the first 20%, two other authors each double-checked 10% of the procedure. Agreement with the first authors as computed with Cohen’s κ were .XX and XX. We resolved disagreements through discussion. We repeated this process once (thus having double-checked 40% in total). In the end, XX articles investigated at least one of the measurement properties outlined in the COSMIN taxonomy of measurement properties and were thus selected. The selection list of the articles can be found at our OSF Page: <https://osf.io/n7z4y/>[[9]](#footnote-10). The full selection process is described in Figure 3.



Placeholder for our flowchart for the to-be-presented Figure 3

**Study 2b – Measurement properties, country and population of origin of social connection measures**

**Methods**

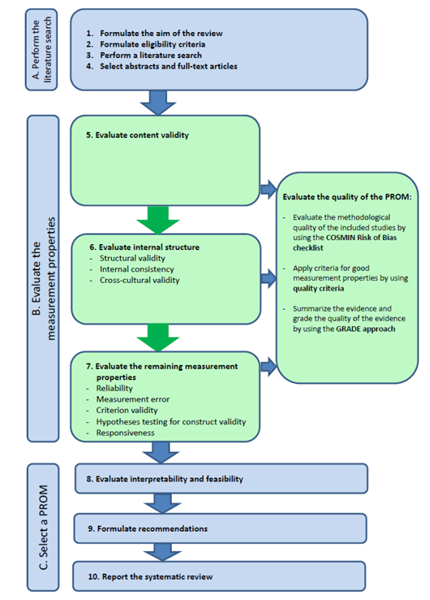
**Evaluation of the measurement properties**

We followed the COSMIN guidelines to evaluate the measurement properties of each measure (see Figure 4). Below we provide a summary of the different steps completed for each measure for replication purposes. We refer the reader to the comprehensive COSMIN manual (Mokkink et al., 2017; Prinsen et al., 2018) for more exhaustive information on how to complete such an evaluation, as the process described below slightly differs across measurement properties.

For each eligible study, we first rated the methodological quality (as *very good*, *adequate*, *doubtful*, or *inadequate*) of that study using the COSMIN Risk of Bias checklist (Mokkink et al., 2017). We then rated the measurement property (as *sufficient*, *insufficient*, or *indeterminate*) reported in the study against the updated criteria for good measurement properties (Prinsen et al., 2018). [Here we will provide a coding example]. This coding process was carried out by [author] and cross-checked by [author]. We resolved conflicts by soliciting the help of an additional coder ([author]).

Once all eligible studies were coded for a given measure, we first summarized the results on each measurement property, either through a) quantitatively pooling the results using meta-analysis, or b) qualitatively summarizing the results, in case quantitatively pooling was not possible. This process allowed us to provide an overall rating (*sufficient*, *insufficient*, or *inconsistent*) for each measurement property. We then accompanied this overall rating with a grading for the quality of evidence (*high*, *moderate*, *low*, or *very low*), using the modified GRADE approached outlined in the COSMIN guidelines. [Here we will provide a coding example]. This coding process was carried out by [author] and cross-checked by [author]. We resolved conflicts by soliciting the help of an additional coder ([author]).

Finally, we described the interpretability and feasibility aspects of each measure (Mokkink et al., 2017; Prinsen et al., 2018; Terwee et al., 2018). Interpretability refers to the degree to which one can assign qualitative meaning to the score obtained to a measure. The interpretability of a measure is notably described by the distribution of scores obtained within a study sample. A measure would typically show great (poor) interpretability if its range of possible scores is wide (narrow). On the other hand, feasibility refers to the ease of application of a measure, in its intended context of use and given various constraints like time or money. The feasibility of a measure is notably described by its number of items. A short measure would typically show greater feasibility than a long measure in most contexts. We extracted the different interpretability and feasibility aspects of each measure using the templates designed by the COSMIN group.



*Figure 4.*

**Evaluation of the country and population of origin**

To determine where the measurements originated and with which population they were developed, we relied on the articles from our search in Study 1a. More specifically, we extracted for each measure a) the country in which the measure was developed, and b) characteristics of the population(s) sampled by the authors of the measure (on age, race, gender, sexuality, religious affiliation, and socioeconomic indicators).

**Results**

**Results on the measurement properties**

Following the PRISMA statement, we reported 1) the characteristics of each measure (constructs being measured, target population, original language, number of items, response options, interpretability, and feasibility) in Table 5; 2) the characteristics of each study (which measure was used, whether the measure was modified, sample size, age, gender, country, language, sampled population) in Table 6; 3) the methodological quality ratings of each study per measurement property (*very good*, *adequate*, *doubtful*, or *inadequate*) along with their results and a rating of these results (*sufficient*, *insufficient*, or *indeterminate*) in Table 7; 4) a summary of findings table containing the overall rating of each measurement property (*sufficient*, *insufficient*, *inconsistent*, or *indeterminate*) along with its quality of evidence grading (*high*, *moderate*, *low*, or *very low*), for each measure, in Table 8.

[We will highlight whether social connection measures have sufficiently good measurement properties or not. For instance, a measure would have sufficiently good reliability property if studies consistently reported an intraclass correlation coefficient or weighted Kappa of at least .70 for that measure].

**Table 5.** Summary of social connection measures’ characteristics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Measure (reference) | Construct(s) | Target population | Original language | Number of items | Response options | Interpretability | Feasibility |
| Measure A (reference) | XX | XX | XX | XX | XX | XX | XX |
| Measure B (reference) | XX | XX | XX | XX | XX | XX | XX |
|  |  |  |  |  |  |  |  |

**Table 6.** Characteristics of studies included in the review

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Study (reference) | Measure | Modified measure | Sample size | *M* Age | *SD* Age | % Female | Country | Language | Sampled population |
| Reference study 1 | XX | XX | XX | XX | XX | XX | XX | XX | XX |
| Reference study 2 | XX | XX | XX | XX | XX | XX | XX | XX | XX |

**Table 7**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |  | |  | |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |  | |  | |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

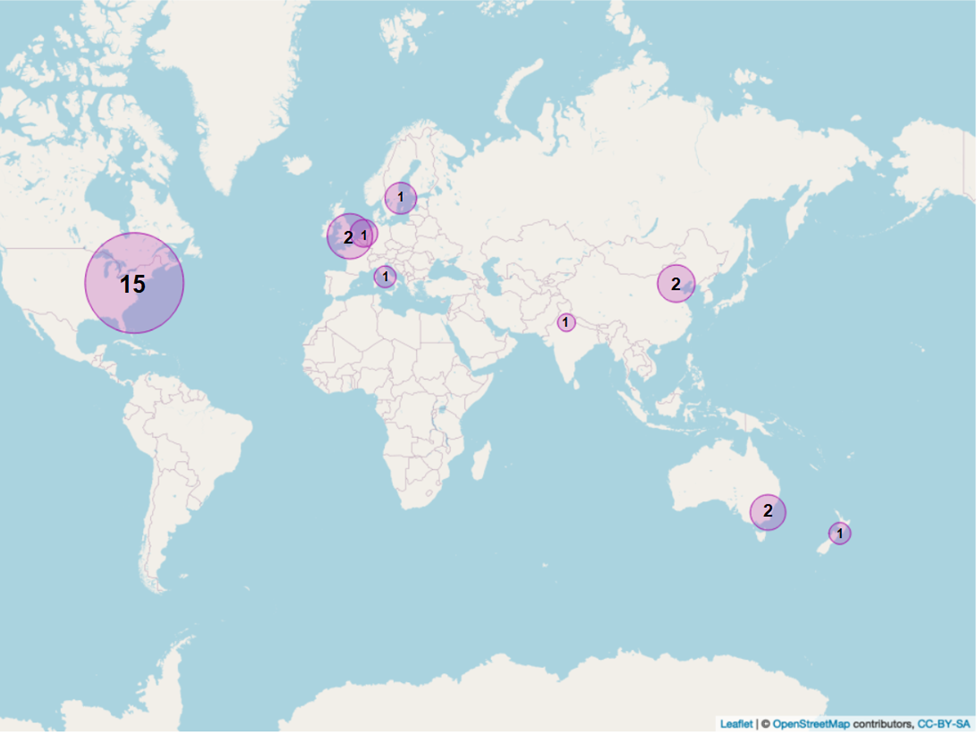
|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | | |  | | |  | | |  |  |  |
|  |  |  |  |  |  |  | |  |  | |  |
|  |  |  |  |  |  |  | |  |  | |  |
|  |  |  |  |  |  |  | |  |  | |  |

**8.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  | |  | |  | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**Results on the country and population of origin**

In Table 9, we report whether age (*M* and *SD*), race, gender, sexuality composition, sample type (e.g., university students, refugees), religious affiliations, and socioeconomic indicators (income, education, employment status) were reported for the sample for which the measure was developed, and if so, what they were. [We will highlight whether the measures were developed with various populations, across various world regions].



*Figure 5*.Country of origin of the XX measures assessing social connection.







|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |  | |  | |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | |  | |  | |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  | |  | | |  | | |  | | |  |  |  |
|  |  |  |  |  |  |  | |  |  | |  |
|  |  |  |  |  |  |  | |  |  | |  |
|  |  |  |  |  |  |  | |  |  | |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  | |  | |  | |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Placeholder for our country of origin, again taken from the CORE Lab (2023).

**Table 9.** *Summary of Social Connection Measures’ Sample Characteristics*

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Measure | *M* Age | *SD* Age | Race Composition | Gender Composition | Sexuality  Composition | Sample Type | Religious Affiliation | Socioeconomic indicators | |
| XX | XX | XX | XX | XX | XX | | XX | XX | XX |

*Table Note*.

**Discussion Study 2b**

[Discussion will be added once analyses have been completed]

**General Discussion**

[General Discussion will be added once all three studies are completed]

**Author Contributions:** Author contributions will be added upon completion of the project.

**Conflict of Interest:** Hans IJzerman is the director of the Annecy Behavioral Science Lab (absl.io), a behavioral science consultancy with an interest in measuring social connection.

Funding: The preparation of this work was partly funded by a Pause Ukraine grant from ANR and a grant from the Fondation Croix-Rouge Française.

**References**

Adetula, A., Forscher, P. S., Basnight-Brown, D., Azouaghe, S., & IJzerman, H. (2022). Psychology should generalize from—Not just to—Africa. *Nature Reviews Psychology*, *1*(7).<https://doi.org/10.1038/s44159-022-00070-y>

American Educational Research Association (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., Norris, S., & Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*(4), 401–406.<https://doi.org/10.1016/j.jclinepi.2010.07.015>

Buecker, S., Mund, M., Chwastek, S., Sostmann, M., & Luhmann, M. (2021). Is loneliness in emerging adults increasing over time? A preregistered cross-temporal meta-analysis and systematic review. *Psychological Bulletin*, *147*(8), 787–805. <https://doi.org/10.1037/bul0000332>

Bugallo-Carrera, C., Dosil-Díaz, C., Anido-Rifón, L., Pacheco-Lorenzo, M., Fernández- Iglesias, M. J., & Gandoy-Crego, M. (2023). A systematic review evaluating loneliness assessment instruments in older adults. *Frontiers in Psychology*, *14*. <https://www.frontiersin.org/articles/10.3389/fpsyg.2023.1101462>

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-Centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science*, *11*(7), 928–937.<https://doi.org/10.1177/1948550620927269>

Cohen, S., Doyle, W. J., Skoner, D. P., Rabin, B. S., & Gwaltney, J. M., Jr. (1997). Social ties and susceptibility to the common cold. *JAMA*, *277*(24), 1940–1944. <https://doi.org/10.1001/jama.1997.03540480040036>

CORE Lab (2023). *A novel, network-based approach to assessing romantic-relationship quality.* PsyArXiv.https://doi.org/10.31234/osf.io/w32dc

Crasta, D., Rogge, R. D., Maniaci, M. R., & Reis, H. T. (2021). Toward an optimized measure of perceived partner responsiveness: Development and validation of the perceived responsiveness and insensitivity scale. *Psychological Assessment*, *33*(4), 338–355.<https://doi.org/10.1037/pas0000986>

Delatorre, M. Z., & Wagner, A. (2020). Marital quality assessment: Reviewing the concept, instruments, and methods. *Marriage & Family Review*, *56*(3), 193–216. <https://doi.org/10.1080/01494929.2020.1712300>

Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465.<https://doi.org/10.1177/2515245920952393>

Fried, E. I. (2017a). The 52 symptoms of major depression: Lack of content overlap among seven common depression scales. *Journal of Affective Disorders*, *208*, 191–197. <https://doi.org/10.1016/j.jad.2016.10.019>

Fried, E. I. (2017b). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, *11*(2), 130–134.<https://doi.org/10.1080/17437199.2017.1306718>

Fried, E. I., & Flake, J. K. (2018). Measurement matters. *APS Observer, 31.* <https://www.psychologicalscience.org/observer/measurement-matters>

Gallup (2022). *The state of social connections methodology report.* Gallup. [https://dataforgood.facebook.com/dfg/resources/state-of-social-connections- methodology-report](https://dataforgood.facebook.com/dfg/resources/state-of-social-connections-%09methodology-report)

Hagland, M. (2019, June 21). *At the AHIP 2019 conference, former surgeon general Murthy speaks of social isolation and its impact.* Healthcare Innovation. [https://www.hcinnovationgroup.com/policy-value-based-care/public- health/article/21085786/at-the-ahip-2019-conference-former-surgeon-general-murthy- speaks-of-social-isolation-and-its-impact](https://www.hcinnovationgroup.com/policy-value-based-care/public-%09health/article/21085786/at-the-ahip-2019-conference-former-surgeon-general-murthy-%09speaks-of-social-isolation-and-its-impact)

Hawkley, L. C., Wroblewski, K., Kaiser, T., Luhmann, M., & Schumm, L. P. (2019). Are U.S. older adults getting lonelier? Age, period, and cohort differences. *Psychology and Aging*, *34*(8), 1144–1157.<https://doi.org/10.1037/pag0000365>

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>

Holt-Lunstad, J. (2018). Why social relationships are important for physical health: A systems approach to understanding and modifying risk and protection. *Annual Review of Psychology*, *69*(1), 437–458.<https://doi.org/10.1146/annurev-psych-122216-011902>

Holt-Lunstad, J. (2021). The major health implications of social connection. *Current Directions in Psychological Science*, *30*(3), 251–259. <https://doi.org/10.1177/0963721421999630>

Holt-Lunstad, J., Robles, T. F., & Sbarra, D. A. (2017). Advancing social connection as a public health priority in the United States. *American Psychologist*, *72*(6), 517–530. <https://doi.org/10.1037/amp0000103>

Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, *10*(2), 227–237.<https://doi.org/10.1177/1745691614568352>

Holt-Lunstad, J., Smith, T. B., & Layton, J. B. (2010). Social relationships and mortality risk: A meta-analytic review. *PLOS Medicine*, *7*(7), e1000316. <https://doi.org/10.1371/journal.pmed.1000316>

Hook, M. K., Gerstein, L. H., Detterich, L., & Gridley, B. (2003). How close are we? Measuring intimacy and examining gender differences. *Journal of Counseling & Development, 81*(4), 462-472. <https://doi.org/10.1002/j.1556-6678.2003.tb00273.x>

IJzerman, H., Dutra, N., Silan, M., Adetula, A., Basnight-Brown, D., & Forscher, P. S. (2021). Psychological science needs the entire globe, part 1. *APS Observer, 34*. <https://www.psychologicalscience.org/observer/global-psych-science>

Keller, H. (2018). Universality claim of attachment theory: Children’s socioemotional development across cultures. *Proceedings of the National Academy of Sciences*, *115*(45), 11414–11419.<https://doi.org/10.1073/pnas.1720325115>

Kocalevent, R.-D., Berg, L., Beutel, M. E., Hinz, A., Zenger, M., Härter, M., Nater, U., & Brähler, E. (2018). Social support in the general population: Standardization of the Oslo social support scale (OSSS-3). *BMC Psychology*, *6*(1), 31. <https://doi.org/10.1186/s40359-018-0249-9>

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for

categorical data. *Biometrics*, *33*(1), 159-174. <https://doi.org/10.2307/2529310>

Maes, M., Qualter, P., Lodder, G. M. A., & Mund, M. (2022). How (not) to measure loneliness: A review of the eight most commonly used scales. *International Journal of Environmental Research and Public Health*, *19*(17), Article 17. <https://doi.org/10.3390/ijerph191710816>

Maggio, L. A., Tannery, N. H., & Kanter, S. L. (2011). Reproducibility of literature search reporting in medical education reviews. *Academic Medicine: Journal of the Association of American Medical Colleges*, *86*(8), 1049–1054. <https://doi.org/10.1097/ACM.0b013e31822221e7>

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*(1), 1.<https://doi.org/10.1186/2046-4053-4-1>

Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of Life Research*, *27*(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, *63*(7), 737–745.<https://doi.org/10.1016/j.jclinepi.2010.02.006>

OpenAI. (n.d.). *GPT-3.5*. Retrieved November 8, 2023, from <https://platform.openai.com/docs/models/gpt-3-5>

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1), 210. <https://doi.org/10.1186/s13643-016-0384-4>

Pantell, M., Rehkopf, D., Jutte, D., Syme, S. L., Balmes, J., & Adler, N. (2013). Social isolation: A predictor of mortality comparable to traditional clinical risk factors. *American Journal of Public Health*, *103*(11), 2056–2062. <https://doi.org/10.2105/AJPH.2013.301261>

Parkerson, G. R., & Gutman, R. A. (2000). Health-related quality of life predictors of survival and hospital utilization. *Health Care Financing Review*, *21*(3), 171–184. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4194681/>

Pomeroy, M. L., Mehrabi, F., Jenkins, E., O’Sullivan, R., Lubben, J., & Cudjoe, T. K. M. (2023). Reflections on measures of social isolation among older adults. *Nature Aging*, 1–2.<https://doi.org/10.1038/s43587-023-00472-4>

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., de Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient- reported outcome measures. *Quality of Life Research*, *27*(5), 1147–1157. <https://doi.org/10.1007/s11136-018-1798-3>

R Core Team (2022). *R: A language and environment for statistical computing.* R Foundation for Statistical computing. [https://www.R-project.org](https://www.r-project.org/)

Rodríguez-Artalejo, F., Guallar-Castillón, P., Herrera, M. C., Otero, C. M., Chiva, M. O., Ochoa, C. C., Banegas, J. R., & Pascual, C. R. (2006). Social network as a predictor of hospital readmission and mortality among older patients with heart failure. *Journal of Cardiac Failure*, *12*(8), 621–627.<https://doi.org/10.1016/j.cardfail.2006.06.471>

Ruan, Y., Reis, H. T., Clark, M. S., Hirsch, J. L., & Bink, B. D. (2020). Can I tell you how I feel? Perceived partner responsiveness encourages emotional expression. *Emotion*, *20*(3), 329–342.<https://doi.org/10.1037/emo0000650>

Rusbult, C. E., Martz, J. M., & Agnew, C. R. (1998). The investment model scale: Measuring commitment level, satisfaction level, quality of alternatives, and investment size. *Personal Relationships, 5*(4), 357-387. [https://doi.org/10.1111/j.1475- 6811.1998.tb00177.x](https://doi.org/10.1111/j.1475-%096811.1998.tb00177.x)

Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, *66*(1), 20–40. <https://doi.org/10.1207/s15327752jpa6601_2>

Scheidecker, G. (2017). *Kindheit, kultur und moralische emotionen: zur sozialisation von furcht und wut im ländlichen Madagaskar.* Transcript Verlag.

Shmotkin, D., Blumstein, T., & Modan, B. (2003). Beyond keeping active: Concomitants of being a volunteer in old-old age. *Psychology and Aging*, *18*(3), 602–607. <https://doi.org/10.1037/0882-7974.18.3.602>

Silan, M., Adetula, A., Basnight-Brown, D., Forscher, P. S., Dutra, N., & IJzerman, H. (2021). Psychological science needs the entire globe, part 2. *APS Observer, 34.* [https://www.psychologicalscience.org/observer/psychological-science-needs-the- entire-globe-part-2](https://www.psychologicalscience.org/observer/psychological-science-needs-the-%09entire-globe-part-2)

Spanier, G. B. (1976). Measuring dyadic adjustment: New scales for assessing the quality of marriage and similar dyads. *Journal of Marriage and Family*, *38*(1), 15–28. <https://doi.org/10.2307/350547>

Taylor, H. O., Cudjoe, T. K., Bu, F., & Lim, M. H. (2023). The state of loneliness and social isolation research: current knowledge and future directions. *BMC public health, 23*(1), 1-3. <https://doi.org/10.1186/s12889-023-15967-3>

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. W. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research*, *18*(8), 1115–1123. <https://doi.org/10.1007/s11136-009-9528-5>

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, *27*(5), 1159–1170.[https://doi.org/10.1007/s11136- 018-1829-0](https://doi.org/10.1007/s11136-%09018-1829-0)

UK Government (2018, October 15). *PM launches government’s first loneliness strategy*. UK Government.[https://www.gov.uk/government/news/pm-launches-governments- first-loneliness-strategy](https://www.gov.uk/government/news/pm-launches-governments-%09first-loneliness-strategy)

Valtorta, N. K., Kanaan, M., Gilbody, S., & Hanratty, B. (2016). Loneliness, social isolation and social relationships: What are we measuring? A novel framework for classifying and comparing tools. *BMJ Open*, *6*(4), e010799. <https://doi.org/10.1136/bmjopen-2015-010799>

Visontay, R., Sunderland, M., Grisham, J., & Slade, T. (2019). Content overlap between youth OCD scales: Heterogeneity among symptoms probed and implications. *Journal of Obsessive-Compulsive and Related Disorders*, *21*, 6–12. <https://doi.org/10.1016/j.jocrd.2018.10.005>

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2017). The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion*, *17*(2), 267–295.<https://doi.org/10.1037/emo0000226>

Zacchilli, T. L., Hendrick, C., & Hendrick, S. S. (2009). The romantic partner conflict scale: A new scale to measure relationship conflict. *Journal of Social and Personal Relationships, 26*(8), 1073-1096. <https://doi.org/10.1177/0265407509347936>

**Appendix A**

Search strings used for the literature search in Study 1

**Structural component of social connection**

**ENGLISH**

**PUBMED:** (("bereave\*" OR "boyfriend" OR "community engagement" OR "community network" OR "divorce" OR "employment status" OR "family characteristics" OR "family network" OR "family size" OR "family structure" OR "friend\*" OR "girlfriend" OR "home alone" OR "human relation\*" OR "interpersonal relation\*" OR "living alone" OR "living arrangement\*" OR "living with other\*" OR "marital status" OR "marriage" OR "married" OR "monogam\*" OR "network size" OR "network structure" OR "parenting" OR "partner status" OR "peer relation\*" OR "phone contact" OR "physical contact" OR "physical isolation" OR "polyam\*" OR "polygam\*" OR "psychosocial factors" OR "religi\*" OR "sexual relation\*" OR "shared living" OR "sibling\*" OR "size of network" OR "social bonding" OR "social contact" OR "social deprivation" OR "social identity" OR "social interaction" OR "social isolation" OR "social life" OR "social network\*" OR "social participation" OR "social relation\*" OR "social structure" OR "social web" OR "spouse" OR "ties" OR "volunteer status" OR "widow\*" OR "social media interaction" OR "social media engagement" OR "social media participation" OR "social media use" OR "internet relation\*" OR "distance relation\*" OR "digital relation\*" OR "internet access" OR "social media access" OR "phone access") AND ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test"))

**PROQUEST:** TI,AB("bereave\*" OR "boyfriend" OR "community engagement" OR "community network" OR "divorce" OR "employment status" OR "family characteristics" OR "family network" OR "family size" OR "family structure" OR "friend\*" OR "girlfriend" OR "home alone" OR "human relation\*" OR "interpersonal relation\*" OR "living alone" OR "living arrangement\*" OR "living with other\*" OR "marital status" OR "marriage" OR "married" OR "monogam\*" OR "network size" OR "network structure" OR "parenting" OR "partner status" OR "peer relation\*" OR "phone contact" OR "physical contact" OR "physical isolation" OR "polyam\*" OR "polygam\*" OR "psychosocial factors" OR "religi\*" OR "sexual relation\*" OR "shared living" OR "sibling\*" OR "size of network" OR "social bonding" OR "social contact" OR "social deprivation" OR "social identity" OR "social interaction" OR "social isolation" OR "social life" OR "social network\*" OR "social participation" OR "social relation\*" OR "social structure" OR "social web" OR "spouse" OR "ties" OR "volunteer status" OR "widow\*" OR "social media interaction" OR "social media engagement" OR "social media participation" OR "social media use" OR "internet relation\*" OR "distance relation\*" OR "digital relation\*" OR "internet access" OR "social media access" OR "phone access") AND TI,AB("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test")

**SCOPUS:** (TITLE-ABS-KEY ("bereave\*" OR "boyfriend" OR "community engagement" OR "community network" OR "divorce" OR "employment status" OR "family characteristics" OR "family network" OR "family size" OR "family structure" OR "friend\*" OR "girlfriend" OR "home alone" OR "human relation\*" OR "interpersonal relation\*" OR "living alone" OR "living arrangement\*" OR "living with other\*" OR "marital status" OR "marriage" OR "married" OR "monogam\*" OR "network size" OR "network structure" OR "parenting" OR "partner status" OR "peer relation\*" OR "phone contact" OR "physical contact" OR "physical isolation" OR "polyam\*" OR "polygam\*" OR "psychosocial factors" OR "religi\*" OR "sexual relation\*" OR "shared living" OR "sibling\*" OR "size of network" OR "social bonding" OR "social contact" OR "social deprivation" OR "social identity" OR "social interaction" OR "social isolation" OR "social life" OR "social network\*" OR "social participation" OR "social relation\*" OR "social structure" OR "social web" OR "spouse" OR "ties" OR "volunteer status" OR "widow\*" OR "social media interaction" OR "social media engagement" OR "social media participation" OR "social media use" OR "internet relation\*" OR "distance relation\*" OR "digital relation\*" OR "internet access" OR "social media access" OR "phone access") AND TITLE-ABS-KEY ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test"))

***Search terms will be translated into the following languages after In Principle Acceptance:***

**UKRAINIAN, RUSSIAN, FRENCH, DUTCH, GERMAN, TURKISH, ARABIC, SPANISH, ITALIAN**

**Functional component of social connection**

**ENGLISH**

**PUBMED:** (("assistance availability" OR "assistance perception" OR "assistance resource" OR "available assistance" OR "available support" OR "belonging assistance" OR "belonging support" OR "close relation\*" OR "emotional assistance" OR "emotional support" OR "family assistance" OR "family support" OR "felt help" OR "felt support" OR "felt understanding" OR "friend\* assistance" OR "friend\* support" OR "friendship" OR "functional assistance" OR "functional support" OR "group assistance" OR "group support" OR "informal assistance" OR "informal support" OR "interpersonal assistance" OR "interpersonal support" OR "lonel\*" OR "mutual understanding" OR "needing assistance" OR "needing support" OR "network\* assistance" OR "network\* support" OR "partner\* responsiveness" OR "peer assistance" OR "peer rejection" OR "peer support" OR "perceived assistance" OR "perceived help" OR "perceived responsiveness" OR "perceived social isolation" OR "perceived support" OR "perceived understanding" OR "perception\* of assistance" OR "perception\* of help" OR "perception\* of support" OR "personal assistance" OR "personal support" OR "psychosocial assistance" OR "psychosocial support" OR "received assistance" OR "received help" OR "received support" OR "relational assistance" OR "relational support" OR "social assistance" OR "social cohesion" OR "social exclusion" OR "social inclusion" OR "social integration" OR "social rejection" OR "social resource\*" OR "social support" OR "subjective assistance" OR "subjective help" OR "subjective social isolation" OR "subjective support" OR "support availability" OR "support group" OR "support perception" OR "support resource" OR "supporting colleagues" OR "supporting friend\*" OR "supporting group" OR "supporting partner\*" OR "supporting relation\*" OR "supportive colleagues" OR "supportive friend\*" OR "supportive group" OR "supportive partner\*" OR "supportive relation\*" OR "work assistance" OR "work relationship\*" OR "work support") AND ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND ("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric"))

**PROQUEST:** TI,AB("assistance availability" OR "assistance perception" OR "assistance resource" OR "available assistance" OR "available support" OR "belonging assistance" OR "belonging support" OR "close relation\*" OR "emotional assistance" OR "emotional support" OR "family assistance" OR "family support" OR "felt help" OR "felt support" OR "felt understanding" OR "friend\* assistance" OR "friend\* support" OR "friendship" OR "functional assistance" OR "functional support" OR "group assistance" OR "group support" OR "informal assistance" OR "informal support" OR "interpersonal assistance" OR "interpersonal support" OR "lonel\*" OR "mutual understanding" OR "needing assistance" OR "needing support" OR "network\* assistance" OR "network\* support" OR "partner\* responsiveness" OR "peer assistance" OR "peer rejection" OR "peer support" OR "perceived assistance" OR "perceived help" OR "perceived responsiveness" OR "perceived social isolation" OR "perceived support" OR "perceived understanding" OR "perception\* of assistance" OR "perception\* of help" OR "perception\* of support" OR "personal assistance" OR "personal support" OR "psychosocial assistance" OR "psychosocial support" OR "received assistance" OR "received help" OR "received support" OR "relational assistance" OR "relational support" OR "social assistance" OR "social cohesion" OR "social exclusion" OR "social inclusion" OR "social integration" OR "social rejection" OR "social resource\*" OR "social support" OR "subjective assistance" OR "subjective help" OR "subjective social isolation" OR "subjective support" OR "support availability" OR "support group" OR "support perception" OR "support resource" OR "supporting colleagues" OR "supporting friend\*" OR "supporting group" OR "supporting partner\*" OR "supporting relation\*" OR "supportive colleagues" OR "supportive friend\*" OR "supportive group" OR "supportive partner\*" OR "supportive relation\*" OR "work assistance" OR "work relationship\*" OR "work support") AND TI,AB("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND TI,AB("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric")

**SCOPUS:** (TITLE-ABS-KEY ("assistance availability" OR "assistance perception" OR "assistance resource" OR "available assistance" OR "available support" OR "belonging assistance" OR "belonging support" OR "close relation\*" OR "emotional assistance" OR "emotional support" OR "family assistance" OR "family support" OR "felt help" OR "felt support" OR "felt understanding" OR "friend\* assistance" OR "friend\* support" OR "friendship" OR "functional assistance" OR "functional support" OR "group assistance" OR "group support" OR "informal assistance" OR "informal support" OR "interpersonal assistance" OR "interpersonal support" OR "lonel\*" OR "mutual understanding" OR "needing assistance" OR "needing support" OR "network\* assistance" OR "network\* support" OR "partner\* responsiveness" OR "peer assistance" OR "peer rejection" OR "peer support" OR "perceived assistance" OR "perceived help" OR "perceived responsiveness" OR "perceived social isolation" OR "perceived support" OR "perceived understanding" OR "perception\* of assistance" OR "perception\* of help" OR "perception\* of support" OR "personal assistance" OR "personal support" OR "psychosocial assistance" OR "psychosocial support" OR "received assistance" OR "received help" OR "received support" OR "relational assistance" OR "relational support" OR "social assistance" OR "social cohesion" OR "social exclusion" OR "social inclusion" OR "social integration" OR "social rejection" OR "social resource\*" OR "social support" OR "subjective assistance" OR "subjective help" OR "subjective social isolation" OR "subjective support" OR "support availability" OR "support group" OR "support perception" OR "support resource" OR "supporting colleagues" OR "supporting friend\*" OR "supporting group" OR "supporting partner\*" OR "supporting relation\*" OR "supportive colleagues" OR "supportive friend\*" OR "supportive group" OR "supportive partner\*" OR "supportive relation\*" OR "work assistance" OR "work relationship\*" OR "work support") AND TITLE-ABS-KEY ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND TITLE-ABS-KEY ("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric"))

***Search terms will be translated into the following languages after In Principle Acceptance:***

**UKRAINIAN, RUSSIAN, FRENCH, DUTCH, GERMAN, TURKISH, ARABIC, SPANISH, ITALIAN**

**Qualitative component of social connection**

**ENGLISH**

**PUBMED:** (("attachment" OR "betrayal" OR "bullying" OR "cheating" OR "close relationship\*" OR "domestic abuse" OR "domestic violence" OR "emotional abuse" OR "emotional violence" OR "family bond\*" OR "family conflict\*" OR "family quality" OR "family relationship\*" OR "friend\* quality" OR "friend\* satisfaction" OR "intimacy" OR "jealous\*" OR "love" OR "marital quality" OR "marital satisfaction" OR "marriage quality" OR "miscommunication" OR "relation\* adjustment" OR "relation\* ambivalence" OR "relation\* commitment" OR "relation\* conflict" OR "relation\* intrusiveness" OR "relation\* quality" OR "relation\* satisfaction" OR "relation\* strain" OR "role\* satisfaction" OR "romantic relationship\*" OR "social conflict" OR "social hostility" OR "social strain" OR "trust") AND ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND ("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric"))

**PROQUEST:** TI,AB("attachment" OR "betrayal" OR "bullying" OR "cheating" OR "close relationship\*" OR "domestic abuse" OR "domestic violence" OR "emotional abuse" OR "emotional violence" OR "family bond\*" OR "family conflict\*" OR "family quality" OR "family relationship\*" OR "friend\* quality" OR "friend\* satisfaction" OR "intimacy" OR "jealous\*" OR "love" OR "marital quality" OR "marital satisfaction" OR "marriage quality" OR "miscommunication" OR "relation\* adjustment" OR "relation\* ambivalence" OR "relation\* commitment" OR "relation\* conflict" OR "relation\* intrusiveness" OR "relation\* quality" OR "relation\* satisfaction" OR "relation\* strain" OR "role\* satisfaction" OR "romantic relationship\*" OR "social conflict" OR "social hostility" OR "social strain" OR "trust") AND TI,AB("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND TI,AB("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric")

**SCOPUS:** (TITLE-ABS-KEY ("attachment" OR "betrayal" OR "bullying" OR "cheating" OR "close relationship\*" OR "domestic abuse" OR "domestic violence" OR "emotional abuse" OR "emotional violence" OR "family bond\*" OR "family conflict\*" OR "family quality" OR "family relationship\*" OR "friend\* quality" OR "friend\* satisfaction" OR "intimacy" OR "jealous\*" OR "love" OR "marital quality" OR "marital satisfaction" OR "marriage quality" OR "miscommunication" OR "relation\* adjustment" OR "relation\* ambivalence" OR "relation\* commitment" OR "relation\* conflict" OR "relation\* intrusiveness" OR "relation\* quality" OR "relation\* satisfaction" OR "relation\* strain" OR "role\* satisfaction" OR "romantic relationship\*" OR "social conflict" OR "social hostility" OR "social strain" OR "trust") AND TITLE-ABS-KEY ("assessment" OR "evaluation" OR "index" OR "instrument" OR "inventory" OR "measure\*" OR "questionnaire" OR "rating" OR "scale\*" OR "self-assessment" OR "self-rated" OR "self-report" OR "survey" OR "test") AND TITLE-ABS-KEY ("CFA" OR "confirmatory factor analysis" OR "cronbach" OR "EFA" OR "exploratory factor analysis" OR "internal consistency" OR "psychometric"))

***Search terms will be translated into the following languages after In Principle Acceptance:***

**UKRAINIAN, RUSSIAN, FRENCH, DUTCH, GERMAN, TURKISH, ARABIC, SPANISH, ITALIAN**

**Appendix B**

Selection criteria for including measures in Study 1

|  |  |
| --- | --- |
| # | Description |
| 1 | The original authors of the measure intended to measure social connection. |
| 2 | The measures were focused on adults (i.e., 18+). |
| 3 | The populations on which the measures were tested were non-clinical. |
| 4 | The measures were self-report measures that target people’s individual attitudes. |
| 5 | The article of the measure was in English, Dutch, Ukrainian, Russian, German, Turkish, Arabic, Spanish, Italian, or French, but the measure could be in any language. |
| 6 | For articles retrieved with the search targeting either the function or qualitative components of social relationships, one goal of the paper in which the measure was presented was to develop/translate/validate the measure. |
| 7 | The full text was available. |
| 8 | The full questionnaire was available. |
| 9 | The study could be published any year. |

**Appendix C**

Template of the prompts we sent to gpt-3.5 in Study 1. We sent one prompt for each item to categorize.

I will share [N] categories for classifying items based on their content:

[CATEGORY 1 NAME: CATEGORY 1 DESCRIPTION]

[CATEGORY 2 NAME: CATEGORY 2 DESCRIPTION]

[CATEGORY 3 NAME: CATEGORY 3 DESCRIPTION]

[CATEGORY 4 NAME: CATEGORY 4 DESCRIPTION]

…

[CATEGORY N NAME: CATEGORY N DESCRIPTION]

Now, classify the following item into the best-fitting category based on its content: [ITEM CONTENT]

**Appendix D**

Search strings used for the literature search in Study 2.

**PUBMED:** (MEASUREMENT NAME OR ABBREVIATED MEASUREMENT NAME) AND (instrumentation[sh] OR methods[sh] OR “Validation Studies”[pt] OR “Comparative Study”[pt] OR “psychometrics”[MeSH] OR psychometr\*[tiab] OR clinimetr\*[tw] OR clinometr\*[tw] OR “outcome assessment (health care)”[MeSH] OR “outcome assessment”[tiab] OR “outcome measure\*”[tw] OR “observer variation”[MeSH] OR “observer variation”[tiab] OR “Health Status Indicators”[MeSH] OR “reproducibility of results”[MeSH] OR reproducib\*[tiab] OR “discriminant analysis”[MeSH] OR reliab\*[tiab] OR unreliab\*[tiab] OR valid\*[tiab] OR “coefficient of variation”[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR “internal consistency”[tiab] OR (cronbach\*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation\*[tiab] OR selection\*[tiab] OR reduction\*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR “precise values”[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab\*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa’s[tiab] OR kappas[tiab] OR repeatab\*[tw] OR ((replicab\*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza\*[tiab] OR generalisa\*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation\*[tiab]) OR discriminative[tiab] OR “known group”[tiab] OR “factor analysis”[tiab] OR “factor analyses”[tiab] OR “factor structure”[tiab] OR “factor structures”[tiab] OR dimension\*[tiab] OR subscale\*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR “item discriminant”[tiab] OR “interscale correlation\*”[tiab] OR error[tiab] OR errors[tiab] OR “individual variability”[tiab] OR “interval variability”[tiab] OR “rate variability”[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR “standard error of measurement”[tiab] OR sensitiv\*[tiab] OR responsive\*[tiab] OR (limit[tiab] AND detection[tiab]) OR “minimal detectable concentration”[tiab] OR interpretab\*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small\*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR “meaningful change”[tiab] OR “ceiling effect”[tiab] OR “floor effect”[tiab] OR “Item response model”[tiab] OR IRT[tiab] OR Rasch[tiab] OR “Differential item functioning”[tiab] OR DIF[tiab] OR “computer adaptive testing”[tiab] OR “item bank”[tiab] OR “cross-cultural equivalence”[tiab]) NOT (“addresses”[Publication Type] OR “biography”[Publication Type] OR “case reports”[Publication Type] OR “comment”[Publication Type] OR “directory”[Publication Type] OR “editorial”[Publication Type] OR “festschrift”[Publication Type] OR “interview”[Publication Type] OR “lectures”[Publication Type] OR “legal cases”[Publication Type] OR “legislation”[Publication Type] OR “letter”[Publication Type] OR “news”[Publication Type] OR “newspaper article”[Publication Type] OR “patient education handout”[Publication Type] OR “popular works”[Publication Type] OR “congresses”[Publication Type] OR “consensus development conference”[Publication Type] OR “consensus development conference, nih”[Publication Type] OR “practice guideline”[Publication Type]) NOT (“animals”[MeSH Terms] NOT “humans”[MeSH Terms])

**PROQUEST:** TI,AB(MEASUREMENT NAME OR ABBREVIATED MEASUREMENT NAME) AND TI,AB(psychometr\* OR clinimetr\* OR clinometr\* OR "outcome assessment" OR "outcome measure\*" OR "observer variation" OR reproducib\* OR reliab\* OR unreliab\* OR valid\* OR "coefficient of variation" OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR "cronbach\* alpha\*" OR "item correlation\*" OR "item selection\*" OR "item reduction\*" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR "test retest" OR "reliab\* test" OR "reliab\* retest" OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa’s OR kappas OR repeatab\* OR "replicab\* measure\*" OR "replicab\* findings" OR "replicab\* result\*" OR "replicab\* test\*" OR "repeated measure\*" OR "repeated findings" OR "repeated result\*" OR "repeated test\*" OR generaliza\* OR generalisa\* OR concordance OR "intraclass correlation\*" OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures" OR dimension\* OR subscale\* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation\*" OR error OR errors OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR "variability values" OR "uncertainty measurement" OR "uncertainty measuring" OR "standard error of measurement" OR sensitiv\* OR responsive\* OR "limit detection" OR "minimal detectable concentration" OR interpretab\* OR "minimal\* important change" OR "minimal\* significant change" OR "minimal\* detectable change" OR "clinical\* important change" OR "clinical\* significant change" OR "clinical\* detectable change" OR "minimal\* important difference" OR "minimal\* significant difference" OR "minimal\* detectable difference" OR "clinical\* important difference" OR "clinical\* significant difference" OR "clinical\* detectable difference" OR "small\* real change" OR "small\* real difference" OR "small\* detectable change" OR "small\* detectable difference" OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")

**SCOPUS:** TITLE-ABS-KEY(MEASUREMENT NAME OR ABBREVIATED MEASUREMENT NAME) AND TITLE-ABS-KEY(psychometr\* OR clinimetr\* OR clinometr\* OR "outcome assessment" OR "outcome measure\*" OR "observer variation" OR reproducib\* OR reliab\* OR unreliab\* OR valid\* OR "coefficient of variation" OR coefficient OR homogeneity OR homogeneous OR "internal consistency" OR "cronbach\* alpha\*" OR "item correlation\*" OR "item selection\*" OR "item reduction\*" OR agreement OR precision OR imprecision OR "precise values" OR test-retest OR "test retest" OR "reliab\* test" OR "reliab\* retest" OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappa’s OR kappas OR repeatab\* OR "replicab\* measure\*" OR "replicab\* findings" OR "replicab\* result\*" OR "replicab\* test\*" OR "repeated measure\*" OR "repeated findings" OR "repeated result\*" OR "repeated test\*" OR generaliza\* OR generalisa\* OR concordance OR "intraclass correlation\*" OR discriminative OR "known group" OR "factor analysis" OR "factor analyses" OR "factor structure" OR "factor structures" OR dimension\* OR subscale\* OR "multitrait scaling analysis" OR "multitrait scaling analyses" OR "item discriminant" OR "interscale correlation\*" OR error OR errors OR "individual variability" OR "interval variability" OR "rate variability" OR "variability analysis" OR "variability values" OR "uncertainty measurement" OR "uncertainty measuring" OR "standard error of measurement" OR sensitiv\* OR responsive\* OR "limit detection" OR "minimal detectable concentration" OR interpretab\* OR "minimal\* important change" OR "minimal\* significant change" OR "minimal\* detectable change" OR "clinical\* important change" OR "clinical\* significant change" OR "clinical\* detectable change" OR "minimal\* important difference" OR "minimal\* significant difference" OR "minimal\* detectable difference" OR "clinical\* important difference" OR "clinical\* significant difference" OR "clinical\* detectable difference" OR "small\* real change" OR "small\* real difference" OR "small\* detectable change" OR "small\* detectable difference" OR "meaningful change" OR "ceiling effect" OR "floor effect" OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR "item bank" OR "cross-cultural equivalence")

1. After having generated a first list of search terms, we used previously published reviews (Bugallo-Carrera et al., 2023; CORE Lab, 2023; Holt-Lunstad et al., 2010; Maes et al., 2022; and Valtorta et al., 2016)

   to validate our search terms in an automated way. We first extracted the paper title and paper DOI associated with each of the 256 articles included in the aforementioned reviews. We then conducted trial searches on Scopus, PubMed, and ProQuest using our generated search terms and exported the results of these searches. Using a Python script, we checked whether each of the 256 articles was found by our trial searches, based on a match between the paper title (paper DOI) of a given article with the paper title (paper DOI) of each result of our trial searches. Our trial searches found 55 (21.48%) of the 256 articles, suggesting that our search terms could be refined. We decided to refine our search terms again using Python scripts. We first downloaded the abstract of the 246 articles for which an abstract was available and checked for the most common monograms, bigrams, trigrams, and quadrigrams across these 246 abstracts, searching for social connection search terms we may have missed. We only selected search terms whose number of occurrences was 10 or above (e.g., social support, marital quality) across the 246 articles and filtered terms that were not relevant for our review (e.g., depression, mortality) leading to a list of 60 search terms that we extracted from these 246 abstracts. Of those 60 search terms, 23 (38%) we had already previously identified and 31 (52%) could not be used because they were too broad. The remaining six search terms (10%) were relevant as a) they produced a reasonable number of new search results and b) were precise enough. We therefore added them to our list of search terms. We then conducted new trial searches using the updated search terms: These new trial searches failed to find target articles missed by the first trial searches. Overall, the results of the data-driven approach we took to improve our search terms suggested that there was little room for improvement. A plausible explanation to these results is the use of development/validation search terms in our search strategy, as the aforementioned systematic reviews did not systematically use these keywords in their search strategy. We shared our Python scripts and a record of that search validation process at <https://osf.io/stmdb/>. [↑](#footnote-ref-2)
2. We are conscious that the differences in search strategies may produce slightly different search results. Of course, the nature of the measures is usually different for structure on the one hand (often measured by single items) and function and quality on the other hand (often measured through compositive indexes and/or Likert scales). One solution could be to remove the development/validation search terms for the function and quality searches. We decided against it after several trial searches. Removing these search terms for the structural component produced many repeated search results. This was not the case for the functional and qualitative components, which likely means that on balance, we would’d lose more relevant results for the functional and qualitative component if we were to select a random subset. By approaching the literature in this way for the functional and qualitative components, we probably lose out on some ad hoc measures. providing a relatively conservative – and thus optimistic – judgment of the measurement quality of the social connection literature. [↑](#footnote-ref-3)
3. Achieving 100% reproducibility of our search strategy is likely not feasible. Instead, we aim to achieve replicability through transparency. By making all the components available, we suspect that a replication of our review – while perhaps not finding exactly the same articles – will likely replicate our main findings. [↑](#footnote-ref-4)
4. These are languages the authorship team speaks fluently. [↑](#footnote-ref-5)
5. We favor “selection criteria” over in/exclusion criteria, as exclusion criteria are invariably the mirror image of inclusion criteria. [↑](#footnote-ref-6)
6. Another useful approach to measure social connection is captured by the Social Relationship Model, which captures target variance, perceiver variance, relationship variance, and error variance. We think that the Social Relationship Model is a useful approach to capture these different sources of variance. At present, we restrict ourselves to a between-person factor model of individual attitudes of social connection. In the medium term, our goal is to switch from between-factor models of individual attitudes of social connection to network attitude models of social connection. In the longer term, we are planning to include other sources of variance (i.e., other people and changes over time) through complex system models. [↑](#footnote-ref-7)
7. For the number of dimensions, we refer to the subscales defined by the authors of the measurement, not to the empirical factors they may have obtained through confirmatory factor analysis. [↑](#footnote-ref-8)
8. We did not write the analysis scripts yet as their content will depend on the data we retrieve from the literature searches. [↑](#footnote-ref-9)
9. We will upload the selection list of articles once we conducted the literature searches. [↑](#footnote-ref-10)